ADVANCES IN CHEMICAL PHYSICS—VOLUME XXVI

I. PRIGOGINE
STUART A. RICE

ADVANCES IN CHEMICAL PHYSICS

NUNC COGNOSCO EX PARTE

THOMAS J. BATA LIBRARY
TRENT UNIVERSITY

Digitized by the Internet Archive
in 2019 with funding from
Kahle/Austin Foundation

https://archive.org/details/advancesinchemic0026unse

ADVANCES IN CHEMICAL PHYSICS

VOLUME XXVI

EDITORIAL BOARD

THOR A. BAK, H. C. Oersted Institute, Copenhagen, Denmark

J. DUCHESNE, University of Liège, Liège, Belgium

H. C. LONGUET-HIGGINS, Department of Machine Intelligence, University of Edinburgh, Edinburgh, Scotland

M. MANDEL, University of Leiden, Leiden, Holland

V. MATHOT, Université Libre de Bruxelles, Brussels, Belgium

P. MAZUR, Institut Lorentz, Leiden, Holland

A. MÜNSTER, Institut für theoretische physikalische Chemie, Frankfurt-am-Main, Germany

S. ONO, Institute of Physics, College of General Education, Tokyo, Japan

B. PULLMAN, Institute de Biologie Physico-Chimique, Université de Paris, Paris, France

J. W. STOUT, The James Franck Institute, University of Chicago, Chicago, Illinois, U.S.A.

G. SZASZ, General Electrical Company, Zurich, Switzerland

M. V. VOLKENSTEIN, Institute of Macromolecular Chemistry, Leningrad, U.S.S.R.

B. H. ZIMM, School of Science and Engineering, University of California at San Diego, La Jolla, California, U.S.A.

Advances in CHEMICAL PHYSICS

EDITED BY

I. PRIGOGINE

University of Brussels,
Brussels, Belgium

AND

STUART A. RICE

Department of Chemistry
and
The James Frank Institute
The University of Chicago
Chicago, Illinois

VOLUME XXVI

AN INTERSCIENCE® PUBLICATION

JOHN WILEY AND SONS

NEW YORK · LONDON · SYDNEY · TORONTO

QD453 .A27 v.26 1974

AN INTERSCIENCE® PUBLICATION

Copyright © 1974, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means, nor transmitted, not translated into a machine language without the written permission of the publisher.

Library of Congress Catalog Card Number: **58–9935**

ISBN 0-471-69931-4

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

INTRODUCTION

In the last decades chemical physics has attracted an ever-increasing amount of interest. The variety of problems, such as those of chemical kinetics, molecular physics, molecular spectroscopy, transport processes, thermodynamics, the study of the state of matter, and the variety of experimental methods used, makes the great development of this field understandable. But the consequence of this breadth of subject matter has been the scattering of the relevant literature in a great number of publications.

Despite this variety and the implicit difficulty of exactly defining the topic of chemical physics, there are a certain number of basic problems that concern the properties of individual molecules and atoms as well as the behavior of statistical ensembles of molecules and atoms. This new series is devoted to this group of problems which are characteristic of modern chemical physics.

As a consequence of the enormous growth in the amount of information to be transmitted, the original papers, as published in the leading scientific journals, have of necessity been made as short as is compatible with a minimum of scientific clarity. They have, therefore, become increasingly difficult to follow for anyone who is not an expert in this specific field. In order to alleviate this situation, numerous publications have recently appeared which are devoted to review articles and which contain a more or less critical survey of the literature in a specific field.

An alternative way to improve the situation, however, is to ask an expert to write a comprehensive article in which he explains his view on a subject freely and without limitation of space. The emphasis in this case would be on the personal ideas of the author. This is the approach that has been attempted in this new series. We hope that as a consequence of this approach, the series may become especially stimulating for new research.

Finally, we hope that the style of this series will develop into something more personal and less academic than what has become the standard scientific style. Such a hope, however, is not likely to be completely realized until a certain degree of maturity has been attained—a process which normally requires a few years.

At present, we intend to publish one volume a year, and occasionally several volumes, but this schedule may be revised in the future.

223834

In order to proceed to a more effective coverage of the different aspects of chemical physics, it has seemed appropriate to form an editorial board. I want to express to them my thanks for their cooperation.

I. PRIGOGINE

CONTRIBUTORS TO VOLUME XXVI

WILLIAM M. GELBART, Department of Chemistry, University of California, Berkeley, California

JUERGEN HINZE, Department of Chemistry, University of Chicago, Chicago, Illinois

E. L. KOSCHMIEDER, College of Engineering and Center for Statistics and Thermodynamics, The University of Texas, Austin, Texas

R. H. MILLER, Department of Astronomy, Institute for Computer Research and Committee on Information Sciences, The University of Chicago, Chicago, Illinois

ELLIOTT W. MONTROLL, Institute for Fundamental Studies, Departments of Physics and Chemistry, University of Rochester, Rochester, New York

J. K. PLATTEN, Department of Thermodynamics, University of Mons, Mons, Belgium

R. S. SCHECHTER, Center for Statistical Mechanics and Thermodynamics, University of Texas, Austin, Texas

M. G. VELARDE, Department de Fisica, Universidad autonoma de Madrid, Canto Blanco, Madrid, Spain

CONTENTS

Depolarized Light Scattering by Simple Fluids 1
By William M. Gelbart

Equilibrium in Stellar Systems 107
By R. H. Miller

Enzyme Cascades and Their Control in Blood Plasma 145
By Elliott W. Montroll

Bénard Convection 177
By E. L. Koschmieder

An Overview of Computational Methods for Large Molecules 213
By Juergen Hinze

The Two-Component Bénard Problem 265
By R. S. Schechter, M. G. Velarde, and J. K. Platten

Author Index 303

Subject Index 313

ADVANCES IN CHEMICAL PHYSICS

VOLUME XXVI

DEPOLARIZED LIGHT SCATTERING BY SIMPLE FLUIDS

WILLIAM M. GELBART

Department of Chemistry, University of California, Berkeley, California

CONTENTS

I. Preface 2
II. Historical Background and Statement of Problem 2
III. Review of Present Theories 6
IV. General Quantum Mechanical Description of Light Scattering by Many-Atom Systems 9
A. Polarized Spectrum 11
B. Depolarized Scattering 16
1. Binary Collision Dynamics 17
2. Three-Body Dynamics 19
3. Higher-Order Effects 21
4. General Discussion of Liquid Case 22
V. Classical (Effective Field) Description of Light Scattering by Many-Atom Systems 24
A. First-Order (Polarized) Inelastic Scattering 27
B. Second-Order (Depolarized) Inelastic Scattering. 28
C. Static Limit: Equivalence to Local Field Analyses 30
D. Relationship to Collision-Induced Pair Scattering 34
VI. Analysis of Data: Appraisal and Discussion 36
A. Isolated Binary Collision Model (Gases) 36
B. Two-Slope Line Shape 39
C. Isolated Binary Collision Model (Liquids) 40
D. Virial (Density) Expansions 43
E. Second-Order Raman Theory 49
F. Lattice Occupancy Models 50
G. Molecular Dynamics "Experiments". 51
VII. Related Optical Processes 55
A. Anisotropic Molecular Fluids and Depolarized Scattering 55
B. Dielectric and Kerr Functions 61
C. Collision-Induced versus Double Scattering in the Gas-Liquid Critical Region 78
D. Summary and Conclusions 92
Appendix A 94
Appendix B 95
Appendix C 97
Appendix D 98
References 101

I. PREFACE

With the development of very intense monochromatic laser light sources and high-resolution interferometric detectors,[1] spectral analysis of the light scattered by *simple fluids* (i.e., *gases and liquids made up of atoms or isotropically polarizable molecules*) has been shown to provide a wealth of information about the correlations of particle motions in these many-body systems. In particular, *polarized* spectra have been studied during the last decade for a wide variety of sample conditions, for example, pure and multicomponent fluids away from and near their critical points, and the theoretical and experimental aspects have been reviewed extensively by many authors.[1,2] The (several orders of magnitude) smaller *de*polarized part (see Section II) of the light scattered by simple fluids has been observed, however, only in the last four years.[3–13] Several competing theories[3–9,11,14–23] have been offered to explain this new effect and considerable controversy (real and apparent) has arisen concerning the most fruitful way to interpret the available data (see Section III). It is the purpose of this review to provide a unified treatment (Sections IV and V) of the depolarized scattering of light by simple gases and liquids, including as special cases all the recent discussions in the literature. We briefly treat also the relationship of this problem to the description of *non*linear polarization effects (e.g., second harmonic generation), critical opalescence, and the definition of dielectric constants and Kerr (namely, electric field induced birefringence) coefficients (Section VII). In this way we hope to emphasize the utility of our model-independent description of the optical properties of simple fluids, and to make precise the connections between, and restrictions of, the various approximate theories. We wish also to comment critically (Section VI) on the several calculations of specific depolarization effects and to appraise thereby what can be learned about molecular motions and interactions in each of these cases.

II. HISTORICAL BACKGROUND AND STATEMENT OF PROBLEM

Over 60 years ago Einstein[24] first calculated the intensity of light scattered by thermal fluctuations in simple liquids. His result relates the observed scattering cross-section to the mean square of the deviation of the local dielectric constant from its equilibrium value: this ensemble average can then be expressed, through thermodynamic[25] arguments, in terms of the isothermal compressibility of the fluid [see (4.33)]. Einstein treated the case of isotropically polarizable scatterers, that is, monatomic and sufficiently symmetric molecular fluids. His conclusions suggested

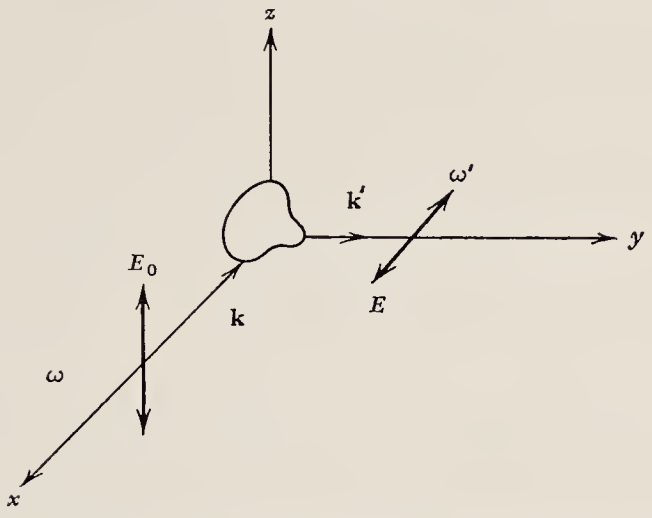


Fig. 1. Right-angle scattering geometry for depolarization experiment.

that the (elastically) scattered light would preserve the polarization of the incident beam and that its intensity would depend linearly on the sample density. Later experiments[26] showed, however, that the light scattered by such fluids includes a nonnegligible *de*polarized component. More recently, advances in laser technology and interferometer design[1] have made it possible to measure the frequency distribution of the depolarized spectrum:[3–6,8–13] the inelastically scattered light is found to be shifted by as much as 100 cm^{-1} to either side of the incident beam and to be characterized by intensities varying as the *square* of the density.

The relevant experimental setup is shown schematically in Fig. 1. Incoming light from an intense (I_0) monochromatic (ω) laser source propagates along the x-axis ($\mathbf{k} = k\hat{x}$) and is linearly polarized in the z-direction ($\mathbf{E}_0 = E_0\hat{z}$); here I_0 is the intensity of the incident beam, ω its frequency, $\mathbf{k}$ its propagation vector, and $\mathbf{E}_0$ (equal to $E_0\hat{z}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega t}$, classically) its electric field amplitude. A detector on the y-axis at some point $\mathbf{R} = R\hat{y}$ far removed from the sample (N fluid particles each characterized by a scalar polarizability α occupying volume V and in thermal equilibrium at temperature T, centered at the origin) measures the intensity of the x-component of the light scattered with frequency ω'. The first experiment of this kind was predicted by Levine and Birnbaum[14a] in 1968 for the noble gases and subsequently performed by McTague and Birnbaum[3] with gaseous samples of Ar and Kr at room temperature and at pressures varying between 20 and 200 atm. Linearly polarized light from a 6328 Å He–Ne laser (~80 mW) was scattered and detected according to the right-angle geometry described above. The inelastic depolarized

spectrum $I^{ZX}(\omega - \omega')$, extended well beyond 50 cm^{-1} to either side of the incident frequency, ω, was characterized by an exponential shape in the wings ($\omega - \omega' \equiv \Omega > 10\ cm^{-1}$), and had an integrated intensity varying as the square of the density for pressures up to 60 atm. In the intervening four years similar experiments have been performed in various laboratories with other monatomic and isotropically polarizable molecular gases and liquids over a wide range of temperatures and pressures.[4–13] We defer discussion of these measurements to Section VI, where we attempt a critical appraisal of the various analyses offered to interpret the existing data.

The total depolarized intensities are found to be thousands of times smaller than the integrated intensities that are observed in the *polarized* case, that is, when the detector is "rotated" to measure the z-component of the scattered light. Also, the frequency distribution $I^{ZZ}(\Omega)$, in the latter case shows a well-known (Rayleigh-Brillouin) triplet structure extending only a few tenths of a cm^{-1} to either side of ω.[2] As we note briefly in Sections IV and V, $I^{ZZ}(\Omega)$, the quasielastic polarized spectrum, can be related directly to the double Fourier transform of the density-density correlation function. The latter can be evaluated from solutions to the linearized hydrodynamic equations and from equilibrium fluctuation theory—this approach allows the polarized line shape to be parameterized in terms of the familiar transport coefficients and thermodynamic derivatives.[27,28] Alternatively, for dilute to moderately dense gases the autocorrelated density can be calculated directly from the appropriate (e.g., Boltzmann, Enskog, etc.) kinetic equation.[29] These important advances have been reviewed already in many excellent articles and we shall not discuss them further. Perhaps it is also appropriate to mention here the several other related topics that are not treated in this review.

1. Depolarized scattering by fluids composed of anisotropic molecules: the $I^{XZ}(\Omega)$ line shapes for gases and liquids of this kind have been related directly to rotational diffusion models and molecular tumbling motions in a large variety of situations.[30–34] Most recently experimental[35–37] and theoretical[38–40] interest has focused on the (small-frequency-shift) angle-dependent fine structure characterizing the depolarized inelastic scattering of light by polyatomic liquids of high viscosity: these new effects have been attributed variously to orientational self-correlations,[38] propagating shear modes,[39] and angular momentum fluctuations;[40] and have been discussed in several critical reviews appearing in a recent conference proceedings.[41] We refer briefly to this situation in Section VII.A, where we consider the role of one- and two-body additive contributions to the

many-body polarizability of a molecular fluid; these respective "mechanisms" ("primary variables") give rise to the sharp and broad components of $I^{XZ}(\Omega)$.

2. Rotational-vibrational Raman spectra: these nonresonant two-photon processes, in which the scattering molecule undergoes a change in its rotational or vibrational state, are well-known and provide a powerful and conventional tool in chemical spectroscopy.[42] From the quantum mechanical discussion presented in Section IV it can be seen that the depolarized inelastic scattering of light by atoms and spherical molecules can be rigorously regarded as a Raman effect involving the *translational* (relative motion) degrees of motion.

3. Fast kinetics and polymer motions: recent light scattering theories of fast chemical reactions[43–46] have been extended to include the important case of macromolecular conformation changes. For example, the depolarized line shape associated with helix-coil transitions has been treated in terms of various detailed models for interconversion dynamics.[47]

4. Nonlinear, multiphoton processes: the exceedingly high intensities (e.g., output fluxes up to thousands of MW/cm^2) from pulsed laser sources in particular have provided a dramatic experimental means for studying nonlinear optical phenomena involving virtually all conceivable aspects of multiphoton spectroscopy. This field has been surveyed many times recently[48–50] and continues to attract increasing research interest. A few questions relevant to the present discussion, for example, *second* harmonic generation in *atomic* fluids, however, have not yet been considered in the literature; we treat this particular example in a separate article[51]—see also Section IV.B.4—as an illustration of the role of electronic distortion, that is, of the breakdown of the isolated-particle-polarizability description.

5. Collision-induced absorption: since the first experimental observation in 1949,[52] a considerable amount of laboratory[53] and theoretical[54] effort has been devoted to studies of the pressure-dependent microwave and infrared absorption coefficients of nonpolar gases. The corresponding rotation-vibration bands have been successfully interpreted as arising from the transient dipole moment induced by the short-range intermolecular forces and/or long-range multipolar fields acting within the collision complexes. Also, *many-body* contributions to the induced moments and cluster dynamics have been considered in particular[55]: these effects have several obvious analogues in our discussion below, but are not pursued further.

In summary, then, the problem to which we *specifically* address ourselves in this review is the following: what is the microscopic basis for the

*de*polarized *in*elastic scattering of light by simple fluids (i.e., what are the precise analogues of molecular anisotropy and internal degrees of freedom); and how do measurements of these and related optical effects allow for new probes of many-body susceptibilities and dynamics in these systems. Before presenting the general descriptions (Sections IV and V) of the relevant optical processes and their application to particular experimental situations (Sections VI and VII), we consider briefly the several approximate theories current in the literature.

III. REVIEW OF PRESENT THEORIES

1. The first series of criticism of the original phenomenological discussion of light scattering by isotropically polarizable particles centered about the neglect by Einstein[24] of fluctuations in the polarizing field acting on the individual scattering volumes. These efforts prompted the work of Fixman[56] and others,[57–59] in which it was shown that anisotropic corrections to the Lorentz local field could in principle account for the small but nonnegligible depolarization of light by monatomic and spherically symmetric molecular fluids. Mazur[60] has provided a general discussion of these effects in his review of the statistical mechanics of the electromagnetic properties of many-body systems: a related treatment is given by Tanaka[61] in which the explicit time dependence of the density fluctuations is included, thereby giving rise to an *in*elastic depolarized spectrum. We consider this approach in a slightly different context in Section V.

2. More recently Frisch and McKenna[62] suggested that one could attribute the depolarized component to the superposition of doubly scattered light waves. Their final expressions for these contributions refer only to the elastic spectrum: more importantly they did not recognize the equivalence, which we discuss below, of the multiple scattering approach to the anisotropic local field discussion of Fixman. Interesting treatments of higher-(than single)order light scattering, and its relationship to the internal field analysis, had been given earlier by Bullough[63] in his papers on the nonadditive contributions of density and concentration fluctuations to the turbidities of multicomponent fluids. Mountain[64] has applied the double scattering results of Frisch and Mckenna to a study of the temperature dependence of the depolarization ratio of simple fluids near the critical region (see, however, Section VII.c. where we treat the general problem of collision induced vs. long range correlation contributions to the depolarized intensity near the gas-liquid critical point).

3. McTague, Fleury, and Dupre,[9] in analyzing their inelastic depolarized spectra of liquid argon, have applied a second-order Raman theory—developed by Stephen[65] to probe short-wavelength collective modes in

the quantum fluid helium II—which is essentially a generalization of the Frisch and McKenna description to include frequency dependence. They thereby calculate a width for the Rayleigh wing which differs by 40% from their observed value. McTague et al. suggest that this discrepancy is due in large part to the factorization approximation introduced for the four-particle density function. In Section V we follow closely the approach of Stephen, whereas in Section VI we consider the implications of the simplifying assumptions involved, in particular those pertaining to the implicit polarization mechanism.

4. The case of depolarized inelastic scattering of light by dilute gases has been treated by Levine and Birnbaum[3,14] as a collision-induced phenomenon and developed in analogy with related work on pressure-dependent infrared absorption[66] and optical birefringence.[67] More explicitly they consider the new effects as arising from the scattering of light by pairs of interacting atoms of (spherically symmetric) molecules: the anisotropic component of the incremental polarizability induced during collision gives rise to the observed depolarization, and the time varying dipole moment associated with the short-lived two-particle complex accounts for the broad frequency spectrum. This description, although certainly appropriate for low enough pressure gases, is inherently model dependent and, as shown in Section IV, must be appreciated as a special case of the general quantum mechanical theory. The pair results are particularly convenient since [see (4.45)] they allow for straightforward moment analyses of the depolarized line shapes: studies of this kind (namely, Refs. 8, 18, and 19) have been used to determine short-range electronic overlap contributions to parameterized forms for diatom anisotropic polarizabilities. See, however, the discussion in Section VI.A.

5. Lewis and Van Kranendonk,[16] proceeding from earlier work on translational adsorption,[68] have extended the isolated binary collision theory to the case of nondilute gases for which the correlations between successive particle encounters can no longer be neglected. Taking the polarizability anisotropy to be proportional to the magnitude of the pair force (thereby including short-range effects), introducing several simplifying assumptions for the collision dynamics, and using hard-sphere expressions to calculate relative speed and angle deflections, they estimate that the predominant contributions to the intercollisional effects come from the correlations between immediately successive encounters, that is, from three-body interactions. As is seen from the discussion in Sections IV and VI these corrections are responsible for a small N^3 dependence in the intensity which is easily observable.

6. Finally, Litovitz and co-workers[11,69] have attempted to carry over the binary collision model to the case of liquids. It is argued that there are

two primary mechanisms for the depolarized light scattering: the first (the dipole-induced-dipole or DID effect) corresponds to fluctuations in the local field acting on a particle due to the moments induced in its neighbors; and the second (electronic overlap or EO) arises from the polarizability distortion due to short-range interaction of overlapping charge distributions. Litovitz et al. assert that a cancellation of the DID effects occurs at liquid densities and that, because even in a multiple collision a molecule has a low probability of strong overlap with more than one neighbor, the EO effect tends to be additive in a pairwise (isolated binary encounter) sense. This approach has been amplified by Shin[20] and others,[13] and is considered critically in Section VI.

7. As is shown throughout the rest of this review the observables corresponding to the depolarization of light by simple fluids can be expressed as a sum of averages over two-, three-, and four-particle distribution functions. Gray and Ralph,[17] McTague, Ellenson, and Hall,[8] and Gelbart[21] have considered these corrections to the two-body (ρ^2) terms using various approximations for the three- and four-point densities. As is already well appreciated in other statistical mechanical contexts, however, these quantities can be calculated "exactly" by the method of molecular dynamics. Berne, Bishop, and Rahman[70] and Alder, Weis, and Strauss[22] have performed computations of this kind for the case of depolarized scattering by atomic fluids at room temperature. We refer to all these studies later in our appraisal in Section VI.

The above (admittedly sketchy and incomplete) outline should serve to introduce the basic problem associated with depolarization studies of simple gases and liquids. More explicitly, we see that our fundamental task is to unravel the simultaneous contributions to the light scattering spectra from many-body *dynamics* on the one hand, and cluster *polarizabilities* on the other. In most of the theoretical analyses mentioned one makes restrictive assumptions about *either* the dynamics (e.g., one neglects three- and four-particle correlations) *or* the polarizabilities (e.g., one argues that short-range, electronic-overlap effects are not involved). Except for the case of a sufficiently dilute gas, however, we shall see that these kinds of approximations can result in misleading interpretations of the nature of the collisions and susceptibilities responsible for the observed depolarization properties. That is, unless the dynamics of the fluid is correctly described (that is, one properly includes triplet and quartet structure, etc.), one cannot proceed with a fit of the optical data to polarizability parameters (e.g., the coefficient and power in the distortion correction λr^{-p} to the dispersional pair anisotropy $6\alpha^2 r^{-6}$). Similarly, unless one carefully considers the many-body and short-range contributions to the cluster susceptibilities, one must be very cautious about

inferring details of the successively higher-order particle correlations. This practical but fundamental problem is pursued throughout the rest of this review.

In Sections IV and V we present first a simple quantum mechanical description of the scattering of light by monatomic fluids, and then a purely classical discussion. Section VI considers these general approaches, and specifically the approximate theories outlined above, in an attempt to examine the usefulness of linear light scattering depolarization studies in elucidating new features of molecular motions and interactions. Finally, in the concluding portion of the review (Section VII) we discuss several optical phenomena, namely, second harmonic generation, critical opalescence, density dependent dielectric and Kerr functions, etc., the study of which raises questions intimately related to those outlined above.

IV. GENERAL QUANTUM MECHANICAL DESCRIPTION OF LIGHT SCATTERING BY MANY-ATOM SYSTEMS

To make this discussion self-contained we begin by briefly reviewing the time-dependent perturbation theory and its application to the interaction of matter and radiation.[71] Consider a system whose Hamiltonian H can be written as

$$H = H_0 + V \tag{4.1}$$

where the zero-order problem is assumed to be solved:

$$H_0 |k\rangle = E_k |k\rangle \tag{4.2}$$

Then the total wave function,

$$|\Psi(t)\rangle = \sum_k c_k(t) e^{-iE_k t/\hbar} |k\rangle \tag{4.3}$$

satisfying

$$i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = [H_0 + V] |\Psi(t)\rangle \tag{4.4}$$

(i.e., describing the system at time t) and obeying the initial condition

$$|\Psi(t = 0)\rangle = |l\rangle \tag{4.5}$$

is given by

$$c_m(t) = \delta_{ml} + \frac{1}{i\hbar} \sum_k \int_0^t d\tau \langle m| V |k\rangle e^{i\omega_{mk}\tau} c_k(\tau) \tag{4.6}$$

where

$$\omega_{mk} \equiv \frac{1}{\hbar} (E_m - E_k) \tag{4.7}$$

(Properly speaking, the lower limit for the τ-integration in (4.6) should be $-\infty$ and V should be "adiabatically switched on" by the damping factor $e^{-\epsilon|t|}$, $\epsilon \to 0^+$; this familiar precaution assures that no spurious time dependences will arise and is implicitly assumed throughout our discussion.) For $V \ll H_0$, (4.6) can be solved by iteration (successive approximations); that is, we set $C_k(t) = \delta_{kl}$ $(= C_k(0))$ on the right-hand side, giving

$$c_m(t) \approx c_m^{(1)}(t) = \delta_{ml} + \frac{1}{i\hbar}\int_0^t d\tau \langle m|\ V\ |l\rangle e^{i\omega_{ml}\tau} \tag{4.8}$$

Substituting $C_k^{(1)}(t)$ for $C_k(t)$ on the right-hand side of (4.6) provides an improved transition amplitude, valid through second order in the perturbation:

$$c_m(t) \approx c_m^{(2)}(t) = c_m^{(1)}(t) + \left(\frac{1}{i\hbar}\right)^2 \sum_k \int_0^t d\tau \int_0^\tau dT \langle m|\ V\ |k\rangle e^{i\omega_{mk}\tau} \langle k|\ V\ |l\rangle e^{i\omega_{kl}T} \tag{4.9}$$

In the case of the interaction of light with, say, a simple (e.g., monatomic) fluid, H_0 is given by

$$H_0 = H_{\text{fluid}} + H_{\text{radiation}} \tag{4.10}$$

where

$$H_{\text{fluid}} = \sum_\alpha \frac{\mathbf{P}_\alpha^{\,2}}{2M_\alpha} + \sum_i \frac{\mathbf{p}_i^{\,2}}{2m} + U(\{\mathbf{x}_i\}, \{\mathbf{R}_\alpha\}) \tag{4.11}$$

α and i running over all the nuclei and electrons; and[71]

$$H_{\text{radiation}} = \tfrac{1}{2}\sum_{\mathbf{k}} \sum_\alpha \hbar c\, |\mathbf{k}|\, (a_{\mathbf{k}\alpha}^\dagger a_{\mathbf{k}\alpha} + a_{\mathbf{k}\alpha}^\dagger a_{\mathbf{k}\alpha}) \tag{4.12}$$

$\mathbf{k}$ and α running over all possible propagation wave vectors $\mathbf{k}$ and polarizations $\mathbf{e}^{(\alpha)}$. $\mathbf{P}_\alpha$ and $\mathbf{p}_i$ are the momentum operators associated with the αth nucleus (mass M_α) and the ith electron (m), $U(\{\mathbf{x}_i\}, \{\mathbf{R}_\alpha\})$ is the total (coulomb) potential energy, and $a_{\mathbf{k}\alpha}$ and $a_{\mathbf{k}\alpha}^\dagger$ are the usual annihilation and creation operators for the $(\mathbf{k}\alpha)$-photon. Finally, the energy of interaction between the fluid and the radiation field is given by[71]

$$V = -\sum_i \frac{e}{mc}\, \mathbf{p}_i \cdot \mathbf{A}(\mathbf{x}_i) + \sum_\alpha \frac{Z_\alpha e}{M_\alpha c}\, \mathbf{P}_\alpha \cdot \mathbf{A}(\mathbf{R}_\alpha) + \sum_i \frac{e^2}{2mc^2}\, \mathbf{A}(\mathbf{x}_i) \cdot \mathbf{A}(\mathbf{x}_i) + \sum_\alpha \frac{Z^2 e^2}{2M_\alpha c^2}\, \mathbf{A}(\mathbf{R}_\alpha) \cdot \mathbf{A}(\mathbf{R}_\alpha) \tag{4.13}$$

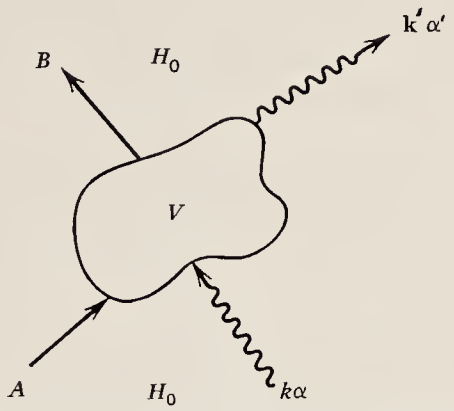


Fig. 2. Schematic rendering of linear (two-photon) light scattering.

Here **A** is the vector potential (satisfying the transversality condition, $\nabla \cdot \mathbf{A} = 0$)

$$\mathbf{A}(\mathbf{r}) = \left(\frac{2\pi\hbar c^2}{\vartheta}\right)^{1/2} \sum_{\bar{\mathbf{k}}} \sum_{\bar{\alpha}} \left(\frac{1}{c\,|\bar{\mathbf{k}}|}\right)^{1/2} [a_{\bar{\mathbf{k}}\bar{\alpha}} \mathbf{e}^{(\bar{\alpha})} e^{i\bar{\mathbf{k}}\cdot\mathbf{r}} + a^{\dagger}_{\bar{\mathbf{k}}\bar{\alpha}}\, \mathbf{e}^{(\bar{\alpha})} e^{-i\bar{\mathbf{k}}\cdot\mathbf{r}}] \quad (4.14)$$

characterizing the quantized electromagnetic field[71]

$$\left(\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \mathbf{B} = \nabla \times \mathbf{A} \quad \text{and} \quad \nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0\right)$$

The normalizing volume ϑ will drop out later. The linear scattering of light by a simple fluid can then be depicted schematically as shown in Fig. 2. $|A\rangle\,|\mathbf{k}\alpha\rangle$ and $|B\rangle\,|\mathbf{k}'\alpha'\rangle$ are eigenfunctions of the zero-order Hamiltonian H_0 and refer to the initial ($|l\rangle$) and final ($|m\rangle$) states of the system, that is, the fluid in state A is subjected to an incoming photon $\mathbf{k}\alpha$ giving rise to an outgoing photon $\mathbf{k}'\alpha'$ and leaving the fluid in state B. The transition amplitude for this process is given by (4.9) with V as in (4.13). We consider first the case of polarized scattering ($\alpha = \alpha'$).

A. Polarized Spectrum

Only the "$\mathbf{A} \cdot \mathbf{A}$" terms in V contribute to $C^{(2)}_{m\leftarrow l}(t)$ in first order—this follows directly from the "raising" and "lowering" properties of the creation and annihilation operators and the orthogonality of the radiation field states. More explicitly, the contribution to $\langle(B;\mathbf{k}'\alpha)|\,V\,|(A;\mathbf{k}\alpha)\rangle$ from the third term in (4.13) can be written as (with $\omega = c\,|\mathbf{k}|$, $\omega' = c\,|\mathbf{k}'|$)

$$\frac{4\pi e^2\hbar}{2m\vartheta(\omega\omega')^{1/2}} \langle B| \sum_i e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_i} |A\rangle \quad (4.15)$$

The matrix element from the fourth term in V is negligible ($m/M_\alpha \ll 1$). If exchange effects are neglected the sum over i can be expressed as $\sum_\alpha \sum_{i\alpha} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_{i\alpha}}$, i_α running over all electrons in the αth atom. Writing $\mathbf{x}_{i_\alpha} \equiv \mathbf{R}_\alpha + (\mathbf{x}_{i_\alpha} - \mathbf{R}_\alpha)$ and noting that atomic dimensions are orders of magnitude smaller than optical wavelengths, the matrix element in (4.15) can be approximated by $\langle B| \sum_\alpha e^{i\mathbf{k}\cdot\mathbf{R}_\alpha} |A\rangle$ where $\mathbf{K} \equiv \mathbf{k} - \mathbf{k}'$ is the momentum transfer corresponding to the difference between the incident and scattered wave vectors. Finally, we use the Born-Oppenheimer separation[72] to describe the initial and final states of the fluid: $|A\rangle = |g\rangle\,|\gamma_g{}^i\rangle$ and $|B\rangle = |g\rangle\,|\gamma_g{}^f\rangle$—here $|g\rangle$ refers to the ground electronic state of the many atom system and $|\gamma_g^{i,f}\rangle$ to the (translational) nuclear wave functions. Thus we have that [compare (4.8)]

$$c^{(t)}_{m\leftarrow l}|\text{“1st order”} = -\frac{2\pi i e^2}{m\vartheta(\omega\omega')^{1/2}} \langle\gamma_g{}^f| \sum_\alpha e^{i\mathbf{K}\cdot\mathbf{R}_\alpha} |\gamma_g{}^i\rangle \int_0^t d\tau e^{i(\omega_{BA}-\Omega)\tau} \tag{4.16}$$

where $\Omega = \omega - \omega'$ is the frequency shift of the scattered light.

Suppose we consider now the second-order contributions to the transition amplitude from the first and second terms in (4.13) (only the first proves significant):

$$c^{(t)}_{m\leftarrow l}|\text{“2nd order”}$$

$$= + \frac{2\pi i e^2}{m^2\vartheta(\omega\omega')^{1/2}} \sum_c \left[\frac{\langle B| \sum_i \mathbf{p}_i\cdot\mathbf{e}^{(\alpha)} e^{-i\mathbf{k}'\cdot\mathbf{x}_i} |C\rangle\langle C| \sum_j \mathbf{p}_j\cdot\mathbf{e}^{(\alpha)} e^{i\mathbf{k}\cdot\mathbf{x}_j} |A\rangle}{E_c - E_A - \hbar\omega} \right.$$

$$\left. + \frac{\langle B| \sum_j \mathbf{p}_j\cdot\mathbf{e}^{(\alpha)} e^{i\mathbf{k}\cdot\mathbf{x}_j} |C\rangle\langle C| \sum_i \mathbf{p}_i\cdot\mathbf{e}^{(\alpha)} e^{-i\mathbf{k}'\cdot\mathbf{x}_i} |A\rangle}{E_c - E_A + \hbar\omega'} \right] \int_0^t d\tau e^{i(\omega_{BA}-\Omega)\tau} \tag{4.17}$$

On writing the $\{|C\rangle\}$ as products of electronic and nuclear factors it is straightforward to show that the only *non*negligible summands in (4.17) are those corresponding to intermediate states of the form $|e\rangle\,|\gamma_e\rangle$ where $e \neq g$ (this is true in general only for sufficiently short-wavelength, e.g., optical, radiation). Neglecting exchange effects as above

$$\left(\sum_i [\mathbf{x}_i] \rightarrow \sum_\alpha \sum_{i_\alpha} [\mathbf{x}_{i_\alpha}]\right)$$

and dropping the translational energy differences $E(\gamma_e) - E(\gamma_g{}^i)$ compared with the electronic origin separations $E_e - E_g$, (4.17) can be expressed in

the form (choosing $\mathbf{e}^{(\alpha)} = \hat{z}$)

$$c^{(t)}_{m\leftarrow l}|\text{"2nd order"} = + \frac{2\pi i e^2}{m^2 \vartheta (\omega\omega')^{1/2}} \langle \gamma_g{}^f | \, c_{zz}(\{\mathbf{R}_\alpha\}) \, | \gamma_g{}^i \rangle \int_0^t d\tau e^{i(\omega_{BA} - \Omega)\tau} \tag{4.18}$$

where

$$c_{zz}(\{\mathbf{R}_\alpha\}) = 2 \sum_\alpha \sum_\beta e^{-i\mathbf{k}'\cdot\mathbf{R}_\alpha} e^{+i\mathbf{k}\cdot\mathbf{R}_\beta} \times \sum_{e\neq g} \frac{\langle g| \sum_{i\alpha} \mathbf{p}_{i_\alpha} \cdot \hat{z} \, |e\rangle\langle e| \sum_{j\beta} \mathbf{p}_{j_\beta} \cdot \hat{z} \, |g\rangle (E_e - E_g)}{(E_e - E_g)^2 - (\hbar\omega)^2} \tag{4.19}$$

In deriving this result we have used the completeness relation

$$\sum_{\gamma_e} |\gamma_e\rangle \langle \gamma_e| = I$$

set $\omega' \approx \omega$ in the energy denominators (the shifts Ω involved are small with respect to optical frequencies), and put $e^{i\mathbf{K}\cdot(\mathbf{x}_{i_\alpha} - \mathbf{R}_\alpha)} \approx 1$ for $\mathbf{K} = +\mathbf{k}$ and $-\mathbf{k}'$.

For $\hbar\omega/(E_e - E_g)$ sufficiently less than unity the sum over $e \neq g$ be approximated by

$$\sum_{e\neq g} \frac{\langle g| \sum_{i_\alpha} \mathbf{p}_{i_\alpha} \cdot \hat{z} \, |e\rangle\langle e| \sum_{j_\beta} \mathbf{p}_{j_\beta} \cdot \hat{z} \, |g\rangle}{(E_e - E_g)} \left[1 + \frac{\hbar^2\omega^2}{(E_e - E_g)^2}\right] \tag{4.20}$$

Using the commutation rules for the electronic momenta and position operators and, in particular, the identity

$$\langle e| \sum_{i_\alpha} \mathbf{p}_{i_\alpha} \, |g\rangle = \frac{im}{\hbar} (E_e - E_g) \langle e| \sum_{i_\alpha} \mathbf{x}_{i_\alpha} \, |g\rangle$$

the contribution to (4.20) from the first term in square brackets can be shown to be equal to $(m/2)\,\delta_{\alpha\beta}$. From (4.18) and (4.19) this is seen to give rise to a contribution to $C^{(t)}_{m\leftarrow l}|\text{"2nd order"}$ which exactly cancels $C^{(t)}_{m\leftarrow l}|\text{"1st order"}$. Thus we have that, through second order in the perturbation, the transition amplitude is given by

$$c^{(t)}_{m\leftarrow e} = + \frac{4\pi i e^2 \hbar^2 \omega^2}{m^2 \vartheta (\omega\omega')^{1/2}} \times \langle \gamma_g{}^f | \sum_\alpha \sum_\beta e^{-i\mathbf{k}'\cdot\mathbf{R}_\alpha} e^{+i\mathbf{k}\cdot\mathbf{R}_\beta} \sum_{e\neq g} \frac{\langle g| \, \mathbf{p}_\alpha \cdot \hat{z} \, |e\rangle\langle e| \, \mathbf{p}_\beta \cdot \hat{z} \, |g\rangle}{(E_e - E_g)^3} \, |\gamma_g{}^i\rangle \times \int_0^t d\tau e^{i(\omega_{BA} - \Omega)\tau} \tag{4.21}$$

($\mathbf{p}_\alpha$ here denotes the total momentum associated with the electrons of α). Suppose we take each electronic wave function to be a product of factors corresponding to the individual atoms (e.g., $|g\rangle = \prod_\alpha |g^{(\alpha)}\rangle$, and similarly for the excited states). (We systematically improve upon this approximation elsewhere[73] where we consider collision induced and double scattering contributions to the polarized spectra of simple fluids. See also our discussion of these effects in Section VII.C. below where we treat the gas liquid critical region.) Then (4.21) can be expressed in the form

$$c^{(t)}_{m\leftarrow l} = -\frac{2\pi i\omega^2\alpha_{zz}}{\vartheta(\omega\omega')^{1/2}}\langle\gamma_g{}^f|\sum_\alpha e^{i\mathbf{k}\cdot\mathbf{R}_\alpha}|\gamma_g{}^i\rangle\int_0^t d\tau e^{i(\omega_{BA}-\Omega)\tau} \tag{4.22}$$

where

$$\alpha_{zz} \equiv 2\hat{e}^2\sum_{e\neq g}\frac{\langle g^{(\alpha)}|\,\mathbf{x}_{(\alpha)}\cdot\hat{z}\,|e^{(\alpha)}\rangle\langle e^{(\alpha)}|\,\mathbf{x}_{(\alpha)}\cdot\hat{z}\,|g^{(\alpha)}\rangle}{(E_e^{(\alpha)} - E_g^{(\alpha)})} \tag{4.23}$$

is the ZZth component of the atomic polarizability ($\alpha_{ZZ} = \alpha_{YY} = \alpha_{XX} = \alpha$). Now, setting $(\omega\omega')^{1/2} \approx \omega$ we have

$$\frac{|c^{(t)}_{m\leftarrow l}|^2}{t} = \frac{8\pi^3\hbar\omega^2\alpha^2}{\vartheta^2}\left|\langle\gamma_g{}^f|\sum_\alpha e^{i\mathbf{k}\cdot\mathbf{R}_\alpha}|\gamma_g{}^i\rangle\right|^2 \otimes \delta(E_B - E_A + \hbar\omega' - \hbar\omega) \tag{4.24}$$

and

$$\text{transition rate}|_{B\leftarrow A} = \int_0^\infty d(\hbar\omega)\rho(\hbar\omega)\frac{|c^{(t)}_{m\leftarrow l}|^2}{t} \tag{4.25}$$

where $\rho_{d\Omega}(\hbar\omega) = \vartheta\omega^2/(2\pi)^3\, d\Omega/\hbar c^3$ is the number of allowed photon states in the energy interval $[\hbar\omega, \hbar\omega + d(\hbar\omega)]$ with propagation vector lying within the solid angle element $d\Omega$. Thus we find for the differential cross-section for polarized scattering

$$\left.\frac{d\sigma}{d\Omega}\right|^{ZZ}_{B\leftarrow A} = \frac{\vartheta}{c\, d\Omega} \otimes \text{transition rate}_{B\leftarrow A} = \alpha^2\left(\frac{\omega}{c}\right)^4\left|\langle\gamma_g{}^f|\sum_\alpha e^{i\mathbf{k}\cdot\mathbf{R}_\alpha}|\gamma_g{}^i\rangle\right|^2 \tag{4.26}$$

or

$$\left.\frac{d\sigma}{d\Omega}\right|^{ZZ} = \sum_{\gamma_g{}^i}\frac{e^{-E(\gamma_g{}^i)/k_BT}}{Z}\sum_{\gamma_g{}^f}\left.\frac{d\sigma}{d\Omega}\right|^{ZZ}_{\gamma_g{}^f\leftarrow\gamma_g{}^i} = \alpha^2\left(\frac{\omega}{c}\right)^4\left\langle\sum_\alpha\sum_\beta e^{i\mathbf{k}\cdot(\mathbf{R}_\alpha-\mathbf{R}_\beta)}\right\rangle \tag{4.27}$$

where, of course, $Z = \sum_{\gamma_g{}^i}\exp[-E(\gamma_g{}^i)/k_gT]$ and the brackets denote a canonical ensemble average. Finally, then, the total (integrated) polarized scattering intensity at a distance R from the sample is given by

$$I^{ZZ} = \left.\frac{d\sigma}{d\Omega}\right|^{ZZ}\frac{I_0}{R^2} = \alpha^2\frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4\bar{\rho}VS(K) \tag{4.28}$$

where $S(K)$ is the familiar static structure factor,

$$S(K) \equiv 1 + \bar{\rho}\int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} g(r) \tag{4.29}$$

with $\bar{\rho} \equiv N/V$ the uniform number density and $g(r)$ the equilibrium radial distribution function.

To determine the *line shape* (frequency distribution) for the polarized scattering we must replace (4.27) by a *restricted* sum over

$$\left.\frac{d\sigma}{d\Omega}\right|_{B\leftarrow A}^{ZZ} (B = \gamma_g{}^f, A = \gamma_g{}^i)$$

including only those initial and final states that "conserve energy," that is, those pairs having energies such that $\hbar\Omega = E(\gamma_g{}^i) - E(\gamma_g{}^f)$. More explicitly,

$$I^{ZZ}(\Omega) = \frac{I_0}{R^2}\sum_{\gamma_g{}^i}\frac{e^{-E(\gamma_g{}^i)/k_BT}}{Z}\sum_{\gamma_g{}^i}\frac{d\sigma}{d\Omega}\Bigg|_{\gamma_g{}^f\leftarrow\gamma_g{}^i}^{ZZ}\delta[\hbar\Omega - (E(\gamma_g{}^i) - E(\gamma_g{}^f))] \tag{4.30}$$

$$= \alpha^2\frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4\frac{1}{2\pi}\int_{-\infty}^{+\infty}dt e^{i\Omega t}\left\langle\sum_\alpha\sum_\beta e^{i\mathbf{k}\cdot\mathbf{R}_\alpha(t)}e^{-i\mathbf{k}\cdot\mathbf{R}_\beta(0)}\right\rangle \tag{4.31}$$

Here we have simply used the Fourier integral form for the Dirac delta function and written $e^{i\mathbf{k}\cdot\mathbf{R}_\alpha(t)} = e^{-iHt/\hbar}(e^{-i\mathbf{k}\cdot\mathbf{R}_\alpha})e^{iHt/\hbar}$ to denote the Heisenberg representation of the exponential operator. Thus

$$I^{ZZ}(\Omega) = \alpha^2\frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4\frac{NS(K,\Omega)}{2\pi} \tag{4.32}$$

where $S(K, \Omega)$ is the double transform of the Van Hove space-time pair correlation function $G^{VH}(rt)$[74] which is defined in the discussion following (5.2). This result agrees exactly with that obtained classically by Komarov and Fisher[75] (see also Pecora[76]) in their application to light scattering theory of the ideas developed earlier for slow neutron spectra of many-body systems.[74] Similarly, (4.28) is reminiscent of the well-known result for elastic X-ray spectra.[77] Noting, however, that for *optical* wavelengths $|\mathbf{K}| \leqslant 10^{5+}\,\mathrm{cm}^{-1}$ implying[78] $S(K) \approx S(0) = k_BT\gamma$ [$\gamma \equiv (\partial\bar{\rho}/\partial P)_T$ [where P and T are the pressure and temperature], we have that the differential cross-section per unit volume (i.e., the Rayleigh ratio) reduces to

$$\left(\frac{\omega}{c}\right)k_BT\gamma\bar{\rho}\alpha^2$$

in agreement with the well-known Einstein relation[1,2]

$$R^{ZZ}_{\text{Einstein}} = \left(\frac{\omega}{c}\right)^4 k_B T \gamma \bar{\rho} \left(\frac{\partial n^2}{4\pi\, \partial \bar{\rho}}\right)^2 \tag{4.33}$$

derived from the phenomenological description of light scattering by density fluctuations. In (4.33) n is the mean refractive index (n^2 the dielectric constant) and, from the Lorentz-Lorenz equation, the factor $(\partial n^2/4\pi\, \partial \bar{\rho})$ can be replaced by the scalar polarizability. We comment further on these results in Section V (see, for example, the last paragraph of Section V.A) where the classical theory is considered in detail.

B. Depolarized Scattering

We consider now the case of "ZX" depolarized scattering.[79] More explicitly we take the incident light to be Z-polarized ($\mathbf{e}^{(\alpha)} = \hat{z}$) and an experimental geometry in which $\mathbf{k}$ is parallel to the x-axis while the detector (and hence $\mathbf{k}'$) lies along the positive y-direction—only those outgoing photons with $\mathbf{e}^{(\alpha')} = \hat{x}$ are measured. In this case the first-order contribution from V to $c_{m\leftarrow i}(t)$ vanishes identically since the matrix element of the "$\mathbf{A}\cdot\mathbf{A}$" terms between the states $\langle B; \mathbf{k}'\mathbf{e}^{(\alpha')}|$ and $|A; \mathbf{k}\mathbf{e}^{(\alpha)}\rangle$ is proportional to $\mathbf{e}^{(\alpha')}\cdot\mathbf{e}^{(\alpha)} = \hat{x}\cdot\hat{z} = 0$. Then, using arguments identical to those discussed in Section IV.A, the transition amplitude can be written (through second order in the perturbation) as

$$c_{m\leftarrow i}(t) = +\frac{2\pi i e^2}{m^2 \vartheta (\omega\omega')^{1/2}} \langle \gamma_g{}^f | A_{XZ}(\{\mathbf{R}_\alpha\}) + B_{ZX}(\{\mathbf{R}_\alpha\}) | \gamma_g{}^i \rangle \int_0^t d\tau e^{i(\omega_{BA}-\Omega)\tau} \tag{4.34}$$

where

$$A_{XZ}(\{\mathbf{R}_\alpha\}) = \sum_{e\neq g} \frac{\langle g| \sum_i (\mathbf{p}_i\cdot\hat{x}) e^{-i\mathbf{k}'\cdot\mathbf{x}_i} |e\rangle\langle e| \sum_j (\mathbf{p}_j\cdot\hat{z}) e^{i\mathbf{k}\cdot\mathbf{x}_j} |g\rangle}{(E_e - E_g) - \hbar\omega} \tag{4.35}$$

and

$$B_{ZX}(\{\mathbf{R}_\alpha\}) = \sum_{e\neq g} \frac{\langle g| \sum_j (\mathbf{p}_j\cdot\hat{z}) e^{i\mathbf{k}\cdot\mathbf{x}_j} |e\rangle\langle e| \sum_i (\mathbf{p}_i\cdot\hat{x}) e^{-i\mathbf{k}'\cdot\mathbf{x}_i} |g\rangle}{(E_e - E_g) + \hbar\omega} \tag{4.36}$$

Since we are interested in the transition *rate* (proportional to the differential scattering cross-section) we need to take the modulus squared of (4.34) and to divide by the time. Then, summing over all final and (Boltzmann weighted) initial states—see Section IV.A—the integrated XZ

intensity is seen to be

$$I_R^{ZX} = \frac{I_0}{R^2}\frac{e^4}{m^4c^4}\langle A_{XZ}^{\dagger}A_{XZ} + B_{ZX}^{\dagger}B_{ZX} + B_{ZX}^{\dagger}A_{XZ} + A_{XZ}^{\dagger}B_{ZX}\rangle \tag{4.37}$$

where the daggers denote hermitian conjugates [e.g., $\langle\gamma_g^f|\, B_{ZX}\,|\gamma_g^i\rangle^* \equiv \langle\gamma_g^i|\, B_{ZX}^{\dagger}\,|\gamma_g^f\rangle$) and the brackets refer as before to a canonical ensemble average.

We wish first to evaluate the ensemble average

$$\int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N\{|A_{XZ}(\{\mathbf{R}_\alpha\}) + B_{ZX}(\{\mathbf{R}_\alpha\})|^2\}P_{eq}^{(N)}(\{\mathbf{R}_\alpha\}) \qquad [\mathrm{I}]$$

appearing in I_R^{ZX}, for the case of a dilute to moderate pressure gas.

1. *Binary Collision Dynamics*

Let, for example, $A_{XZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2)$ be the form assumed by the many-body adiabatic (that is, it depends parametrically on all positions $\{\mathbf{R}_\alpha\} \equiv \mathbf{R}_1, \ldots, \mathbf{R}_N$) quantity A_{XZ} for the particular N-particle configurations for which particles 1 and 2 are close enough to interact with each other, while all others are isolated from one another. As shown in Appendix C, for optical wavelengths and $(\hbar\omega/(E_e - E_g))$ sufficiently less than unity, $A_{XZ}^{(12)}$ can be written as

$$A_{XZ}^{(12)} = \frac{\omega}{2\hbar}\left(\frac{m}{e}\right)^2 e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}^{(12)}}[\hbar\omega\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2) + 2e^2Q_{XZ}(\mathbf{R}_1\mathbf{R}_2)] \tag{4.38}$$

Here $\mathbf{R}_{(12)}$ denotes a vector drawn to the c.m. of the interacting pair of atoms,

$$\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2) \equiv 2e^2\sum_{e\neq g}\frac{\langle g^{(12)}|\,[\mathbf{x}^{(12)}]_X\,|e^{(12)}\rangle\langle e^{(12)}|\,[\mathbf{x}^{(12)}]_Z\,|g^{(12)}\rangle}{E_{e^{(12)}} - E_{g^{(12)}}} \tag{4.39}$$

is the XZ component of the adiabatic polarizability of the pair, and

$$Q_{XZ}(\mathbf{R}_1\mathbf{R}_2) = \langle g^{(12)}|\,[\mathbf{x}^{(12)}]_X[\mathbf{x}^{(12)}]_Z\,|g^{(12)}\rangle \tag{4.40}$$

is the XZ component of the pair's "quadrupole moment." In the above, $\mathbf{x}^{(12)}$ denotes a sum over the position vectors of the electrons associated with atoms 1 and 2, and $|g^{(12)}\rangle$ and $|e^{(12)}\rangle$ the ground and excited adiabatic electronic states of the diatom, with $E_{g(12)}$ and $E_{e(12)}$ the corresponding origins. Similar expressions obtain for $A_{XZ}^{(13)}$, $A_{XZ}^{(23)}$, and so on, that is, for all the distinct pairs of atoms.

Consider all possible configurations of the N atoms. If the system is sufficiently dilute, then the only configurations that contribute non-negligibly to the ensemble average [I] are those in which no atom interacts

with more than one other; that is, each atom is either isolated or interacts with only one other atom. This situation is referred to as the case of *isolated pair* interactions. Now, it should be obvious from the above that in this limit the ensemble average [I] assumes the form

$$\int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \left\{ \left| \sum_{ij} A_{XZ}^{(ij)}(\mathbf{R}_i\mathbf{R}_j) + \sum_{kl} B_{ZX}^{(kl)}(\mathbf{R}_k\mathbf{R}_l) \right|^2 \right\} P^{(N)}(\mathbf{R}_1, \ldots, \mathbf{R}_N) \quad [\mathrm{I}']$$

Here the sums $\sum_{ij}$ and $\sum_{kl}$ each run over all the distinct pairs of particles. We multiply out the product of sums in the first of the four terms in $\{\cdots\}$ and consider the contributions to the ensemble average from the "squared terms," for example,

$$\int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \{A_{XZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2) A_{XZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2)\} P^{(N)}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$$

$$= \int d\mathbf{R}_1 \int d\mathbf{R}_N \, |A_{XZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2)|^2 \, P^{(2)}(\mathbf{R}_1\mathbf{R}_2) \quad (4.41)$$

There are $N(N-1)/2!$ identical contributions of this kind. All the "cross terms" arising in the product of sums vanish identically in the case of isolated pair interactions. The remaining three products ($B_{ZX}{}^{\dagger} B_{ZX}$, $B_{ZX}{}^{\dagger} A_{XZ}$, and $B_{ZX} A_{XZ}{}^{\dagger}$) in $\{\cdots\}$ are simplified in a similar way, and involve

$$B_{ZX}^{(kl)} = \frac{\omega}{2\hbar}\left(\frac{m}{e}\right)^2 e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{(kl)}} [\hbar\omega\alpha_{XZ}(\mathbf{R}_k\mathbf{R}_l) - 2e^2 Q_{XZ}(\mathbf{R}_k\mathbf{R}_l)] \quad (4.42)$$

Adding together all these contributions, and recalling that $A_{XZ}{}^{\dagger} = (A_{XZ})^*$, $B_{ZX}{}^{\dagger} = (B_{ZX})^*$, we have that [I'] can be written as

$$\omega^4 \left(\frac{m}{e}\right)^4 \frac{N(N-1)}{2!} \iint d\mathbf{R}_1 \, d\mathbf{R}_2 [\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)]^2 P^{(2)}(\mathbf{R}_1\mathbf{R}_2) \quad (4.43)$$

Thus, in the limit of isolated binary collision dynamics,

$$I^{ZX} = \frac{I_0}{R^2}\frac{e^4}{m^4c^4}[I] = \frac{I_0}{R^2}\frac{\omega^4}{c^4}\tfrac{1}{2}\rho^2 V \int d\mathbf{r}[\alpha_{XZ}(\mathbf{r})]^2 g(r) \quad (4.44)$$

Noting that $\alpha_{XZ}(\mathbf{r}) \to 2\alpha_0{}^2 T_{XZ}{}^d(\mathbf{r})$ in the induced dipole limit—here α_0 is the atomic scalar polarizability and $\mathsf{T}(\mathbf{r}) \equiv 3\hat{r}\hat{r} - \mathsf{I}/r^3$—(4.44) will be seen to agree with the integrated depolarized intensity obtained classically using an effective (internal) field analysis [see Section V and, in particular, the pair interaction limit of equation (2) in Ref. 21].

As mentioned in Section IV.A, the *in*elastic scattering spectrum, $I^{ZX}(\Omega)$ is obtained by restricting the cross-section sum over initial and final states to those that conserve energy—the delta function constraint gives rise

here to a (time) autocorrelated pair polarizability and the depolarized line shape is given by

$$I^{ZX}(\Omega) = \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \frac{V}{4\pi} \int_{-\infty}^{+\infty} dt e^{i\Omega t} \langle \alpha_{XZ}[\mathbf{r}(t)]\alpha_{XZ}[\mathbf{r}(0)]\rangle \qquad (4.45)$$

Equation (4.45) is *identical* to the ZX-depolarized inelastic spectrum which would be calculated for a sample of anisotropic (α), e.g., homonuclear diatomic molecules for which only "self" correlations are considered.[33] McTague and Birnbaum[3] developed the isolated binary encounter description of collision-induced scattering by suggesting that the formalism for nonspherical molecules be carried over to spherical particles by simply regarding each interacting pair as an axially symmetric molecule whose polarizability is a function of the interparticle separation. Accordingly, for dilute gases, they can write down[3b] immediately a result identical to (4.45). In Section V we find that the classical (effective field) description of the scattering of light by a simple fluid leads to a similar depolarized spectrum except that the pair polarizabilities involved are to be calculated in the induced dipole approximation. Recently many investigators[14,18–20] have used line shape expressions similar to (4.45) to test parameterized models for $\alpha_{XZ}(\mathbf{r})$ which include short-range electronic overlap corrections to the long-range dispersion form. As we point out in Section VI, however, starting with this result implicitly neglects the contributions to the ZX scattering from three- and four-particle correlations (see also Ref. 21).

2. *Three-body Dynamics*

For less dilute gases the ensemble average [I] includes important contributions as well from N-particle configurations for which *three's* of atoms interact simultaneously. (Generalization to still higher-order correlations is entirely straightforward.) We proceed as in the previous section except that now, in addition to the $N(N-1)/2!$ pair terms

$$\sum_{i<j}\sum A_{XZ}^{(ij)} \qquad (4.46)$$

we must explicitly take into account the $N(N-1)(N-2)/3!$ irreducible triplet terms

$$\sum_i \sum_j \sum_k {}^{\text{irr}}A_{XZ}^{(ijk)} \qquad (4.47)$$

${}^{\text{irr}}A_{XZ}^{(ijk)}$ is simply related to

$$A_{XZ}^{(ijk)} = \frac{\omega}{2\hbar}\left(\frac{m}{e}\right)^2 e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}(ijk)} [\hbar\omega\alpha_{XZ}(\mathbf{R}_i\mathbf{R}_j\mathbf{R}_k) + 2e^2 Q_{XZ}(\mathbf{R}_i\mathbf{R}_j\mathbf{R}_k)] \qquad (4.48)$$

the form of A_{XZ} for the N-particle configurations for which particles i, j, and k are all close enough to be interacting with one another, while

all others are isolated. ${}^{\text{irr}}A_{XZ}^{(ijk)}$ is defined, in fact, by

$$ {}^{\text{irr}}A_{XZ}^{(ijk)} \equiv A_{XZ}^{(ijk)} - A_{XZ}^{(ij)} - A_{XZ}^{(ik)} - A_{XZ}^{(jk)} \tag{4.49} $$

That is, ${}^{\text{irr}}A_{XZ}^{(ijk)}$ is the *irreducible* triplet part of the adiabatic triatom polarizability anisotropy.

Writing A_{XZ} in [I] as a sum of (4.46) and (4.47) and similarly for B_{ZX} we can proceed to evaluate the three-body corrections. The ensemble average [I] can be written as

$$ \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \Bigg\{ \Bigg[\sum_{i<j} \sum A_{XZ}^{(ij)}(\mathbf{R}_i\mathbf{R}_j) + \sum_{l} \sum_{m} \sum_{n} {}^{\text{irr}}A_{XZ}^{(lmn)}(\mathbf{R}_l\mathbf{R}_m\mathbf{R}_n) \Bigg] $$
$$ \otimes \Bigg[\sum_{i'<j'} \sum A_{XZ}^{(i'j')}(\mathbf{R}_{i'}\mathbf{R}_{j'}) + \sum_{l'} \sum_{m'} \sum_{n'} \cdots \Bigg]^{*} + \cdots \Bigg\} P^{(N)}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N) \tag{I''} $$

$\sum_{ij}$ runs over all $N(N-1)/2!$ distinct pairs, and $\sum_{lmn}$ over all $N(N-1)\times(N-2)/3!$ clusters of three. The "squared terms" arising from the product $\sum_{ij}\sum_{i'j'}$ give results identical to those obtained in the preceding section for the case of isolated pair interactions. Now, however, the contributions from the "cross terms" no longer vanish. Consider in particular those involving the position vectors of three particles only, for example,

$$ \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \{[A_{XZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2)][A_{XZ}^{(13)}(\mathbf{R}_1\mathbf{R}_3)]^{\dagger} \cdots\} P^{(N)}(\mathbf{R}_1, \ldots, \mathbf{R}_N) $$
$$ = \omega^4 \left(\frac{m}{e}\right)^4 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 \alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_3) P^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) \tag{4.50} $$

There are $2[N(N-1)/2!](N-2)$ identical contributions of this kind. The cross terms involving four different position vectors give zero—they don't begin to contribute until the next step, that is, where we would explicitly allow for *four* particles to interact simultaneously. The contributions from the terms in the product $\sum_{lmn}\sum_{l'm'n'}$ are each of the form

$$ \omega^4 \left(\frac{u}{e}\right)^4 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 [{}^{\text{irr}}\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)]^2 P^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) \tag{4.51} $$

and there are $N(N-1)(N-2)/3!$ of them; again, only the "squared" terms give nonzero results. Finally, using the same arguments as above, the terms arising from the products $\sum_{ij}\sum_{l'm'n'}$ and $\sum_{lmn}\sum_{i'j'}$ are seen to give

$$ 2\,\frac{N(N-1)}{2!}(N-2)\omega^4 \left(\frac{m}{e}\right)^4 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 $$
$$ \times \alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)\, {}^{\text{irr}}\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) P^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) \tag{4.52} $$

Thus for simple gases—sufficiently dilute so that no more than three particles are correlated simultaneously—we can write for the integrated depolarized scattering intensity:

$$I^{ZX} = \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \frac{1}{2}\Big[\rho^2 \iint d\mathbf{R}_1\, d\mathbf{R}_2 [\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)]^2 g(\mathbf{R}_1\mathbf{R}_2)$$
$$+ 2\rho^3 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 [\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)][\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_3)] g^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$$
$$+ \rho^3 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 \Big\{ \frac{[{}^{\text{irr}}\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)]^2}{3}$$
$$+ 2[\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)][{}^{\text{irr}}\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)] \Big\} g^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) \Big] \tag{4.53}$$

Note that only upon dropping the last two terms involving the irreducible triplet polarizability, and writing

$$\alpha_{XZ}(\mathbf{r}) \to 2\alpha_0{}^2 T_{XZ}{}^d(\mathbf{r}) = 2\alpha_0{}^2(3\hat{r}\hat{r} - \mathrm{I})_{XZ}/r^3$$

is (4.53) seen to be identical to the corresponding virial expansion obtained classically by the effective field analysis [compare Part V, and equation (2) of Ref. 21]. The surviving three-body average has been evaluated within the induced dipole (DID) limit for the pair anisotropy and for various choices of intermolecular potentials;[8,16,17,21,22,64,70] these calculations are discussed in Section VI.

3. *Higher-Order Effects*

It should be obvious that one can proceed as above to determine the contributions to the depolarized scattering intensity from quartet and successively higher correlations. It is not particularly instructive to do so, however, since the available experimental data are not sufficiently thorough, and the higher-order ensemble averages become increasingly difficult to calculate. Even for the three-body corrections derived above it is often necessary to employ triplet-density superposition approximations of questionable quantitative accuracy.[8,17,21] Also, one is constrained to write out the distribution functions in (4.53) as power series in ρ since the laboratory measurements of depolarization are perforce expressed in terms of density (rather than virial) expansions. Most importantly, very little is considered to be known about the irreducible polarizability components of third [e.g., ${}^{\text{irr}}\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$] and higher order. In fact, there is still considerable controversy (see Section VI) concerning the proper form of the *pair* anisotropies, $\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)$; hardly any ab initio calculations have been

reported to date on any of the appropriate di-(inert-) atoms or "round" molecule complexes.*

In Section VII.B we argue that the triplet incremental polarizability components ${}^{\text{irr}}\alpha_{\alpha\beta}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)(\alpha, \beta = x, y, z)$ can be expressed as a symmetric sum of products of the pair functions discussed above. It is shown in particular that such a "superposition approximation" for ${}^{\text{irr}}\alpha_{ZZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$ [see equation (7.31)] leads to a self-consistent generalization of the Kirkwood dielectric theory which explicitly includes (many-body) short-range electronic distortion effects as well as the dipole-induced dipole contributions.

4. *General Discussion of Liquid Case*

The quantum mechanical theory described earlier can readily be applied to *high*-pressure gases, and *liquids*, by simply avoiding the virial series evaluation of the ensemble average, [I]. Rather than make any assumptions about the *dynamics*, for example, the relative importance of one-, two-, and three-body correlations, etc., we can suppose only that the (adiabatic) quantities A_{XZ} and B_{ZX} given by (4.35) and (4.36) are pairwise additive. More explicitly, we write $A_{XZ}(\{\mathbf{R}_\alpha\}) \approx \sum_{i<j}\sum A_{XZ}^{(ij)}(\mathbf{R}_i\mathbf{R}_j)$ where $A_{XZ}^{(ij)}$, identical to the function introduced in (4.38), is the form assumed by A_{XZ} for the case where only particles i and j are present. Then [I] $\rightarrow$ [I'] and it is straightforward to show for arbitrary density [note that $P^{(N)}(R_1, \ldots, \mathbf{R}_N)$ now assigns *non*negligible weights—characteristic of the high-pressure gas, or liquid, in question—to the N-particle configurations involving *many*-body effects]:

$$I^{ZX} = \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \frac{1}{2}\Bigg[\rho^2 \iint d\mathbf{R}_1\, d\mathbf{R}_2[\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)]^2 g(\mathbf{R}_1\mathbf{R}_2)$$
$$+ 2\rho^3 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3[\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)][\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_3)]g^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$$
$$+ \frac{\rho^4}{2} \iiiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3\, d\mathbf{R}_4[\alpha_{XZ}(\mathbf{R}_1\mathbf{R}_2)\alpha_{XZ}(\mathbf{R}_3\mathbf{R}_4)]g^{(4)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3\mathbf{R}_4)\Bigg] \tag{4.54}$$

Without our having made any assumptions about the dynamics, the highest-order distribution function that appears above is that corresponding to four-particle correlations. This is simply because of (*1*) the supposition of pairwise additivity for the generalized susceptibilities A_{XZ} and B_{ZX}; and (*2*) the use of second-order perturbation theory. It is important to note that starting points equivalent to (*1*) and (*2*) are implicitly invoked

* See, however, the recent SCF results provided by O'Brien, Gutschick, McKoy, and McTague[80] and our discussion at the end of Sections VI.D and III.G, and VII.B.

as well in the classical effective field analysis of ZX-scattering where neglect of electronic overlap effects allows a local electric amplitude to be defined and the resulting Maxwell equations are solved by successive approximations only through second order—this point is discussed further in Sections V and VII. It is thus not surprising that the expression obtained in the latter case (compare equation (2) of Ref. 21) for the integrated intensity I^{ZX} is identical to (4.54) except for the adiabatic anisotropy $\alpha_{XZ}(\mathbf{r})$ being replaced by its long-range dispersion form. An entirely analogous situation obtains in the case of the *polarized* cross-section I^{ZZ} for simple gases and liquids; upon lifting the usual restriction of a *one*-body-additive polarization, corrections to the familiar structure factor can be derived directly from two-, three-, and four-particle correlations.

Recently, molecular dynamics calculations[22,70] have been performed which address themselves specifically to the "exact" numerical evaluation of the ensemble averages appearing in the depolarization, (4.54). These computations correspond to an induced dipole approximation for the pair polarizabilities (with accompanying neglect of higher-order irreducible components) and yet provide considerable insight into the relationships between the many-particle dynamics and the measured cross-sections and line shapes. In particular the *qualitative* form of the density dependence of the integrated intensity $[I^{ZX}]$ is correctly reproduced, as well as the break in slope observed in the inelastic spectrum $[\ln I^{ZX}(\Omega)]$. The significance of these results is discussed in Section VI with special emphasis on the role of shorter-range contributions to the cluster susceptibilities (namely, those included in the exact anisotropies above).

It should be mentioned here that the above analysis can be applied directly to probing *hyper*polarizabilities and many-body effects in second harmonic generation (SHG) and related *non*linear optical properties of dense isotropic media. The use of *third*-order perturbation theory to describe the relevant *three*-photon scattering processes leads to transition rates involving the *cubic* (as opposed to the quadratic term considered above) susceptibility $\boldsymbol{\beta}$ defined by

$$\boldsymbol{\mu}|_{\text{Cluster}} \equiv \boldsymbol{\beta}|_{\text{Cluster}} : \mathbf{EE} \tag{4.55}$$

with $\boldsymbol{\mu}$ the dipole moment induced in a many-particle grouping by the incident electric field amplitude $\mathbf{E}$. Since $\boldsymbol{\beta}$ vanishes identically for centrosymmetric clusters, the first contribution to the virial expansion for the nonlinear (e.g., SHG) cross-section arises from three-body effects. (The situation is similar for the case of translational absorption in simple fluids.) The quantum mechanical $\beta_{ijk}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$ includes, for example, the fourth-order *atomic* polarizabilities discussed by Kielich[81] *plus* the corrections which correspond to deviations from the usual induced multipole

description. Thus, as in the linear, depolarized scattering treated above, it is important to go beyond the assumption of nonoverlapping electronic charge distributions and isolated-particle polarizabilities in calculating these collision-induced optical properties: we consider the nonlinear case in a separate article.[51]

Finally we remark that the simple physical arguments used in Sections IV.B.1 and IV.B.2 above to derive a virial series for I^{ZX} are essentially the same as a general method for generating density expansions of ensemble averages which has been discussed most recently by Ramshaw.[82] More explicitly, it has been shown that a general canonical statistical mechanical average $F(N, V, T)$ can be written, in the thermodynamic limit ($N, V \to \infty$, $N/V \equiv \rho = \text{constant}$), as

$$F(\rho, T) = \sum_{k=0}^{\infty} A_k(T)\rho^k \tag{4.56}$$

with specified radius of convergence, where

$$A_k(T) \equiv (-1)^k \lim_{V\to\infty} \left[V^k \sum_{j=0}^{k} \frac{(-1)^j}{j!\,(k-j)!} F(j, V, T) \right] \tag{4.57}$$

Here $F(j, V, T)$ is the form of the ensemble average in question for a system containing only j molecules. In our case

$$F(N, V, T) \equiv \langle\{F\}\rangle = \langle\{A_{XZ}A_{XZ}^{\dagger} + \cdots\}\rangle = [I]$$

the integral defined in the very beginning of this section. It is then altogether straightforward to evaluate $A_2(T)$ and $A_3(T)[A_0(T) = 0 = A_1(T)]$ in terms of the irreducible pair and triplet polarizabilities and to derive in this way a virial series identical to our (4.53).

V. CLASSICAL (EFFECTIVE FIELD) DESCRIPTION OF LIGHT SCATTERING BY MANY-ATOM SYSTEMS

Traditionally, the classical microscopic (statistical mechanical) theory of light scattering by simple fluids requires an evaluation of the effective electric vector (i.e., the field amplitude) at an arbitrary point due to the incident beam and its polarizing effect on the surrounding volume. Consistent with the well-known and successful dielectric theory of Kirkwood,[83] Yvon,[84] and others,[85] it is assumed that the optical properties of the many-atom system can be described in the dipole-induced-dipole (DID) approximation. The use of a multipole expansion, however, depends implicitly on the charge distributions being nonoverlapping—the quantum mechanical theory of the previous section does not require any such restriction and is parameterized instead in terms of *exact* polarizabilities. The implications of this starting point are discussed in detail in Section VI.

We consider again a system of N atoms of spherical molecules (each characterized by a scalar polarizability α) enclosed in a volume V and in thermal equilibrium at a temperature T. In the DID approximation the effective (local) electric field $\mathbf{E}(\mathbf{r}, t)$ at a point $\mathbf{r} \in V$ and time t consists of the incident field $\mathbf{E}_0(\mathbf{r}, t)$ plus the contribution from all the induced dipoles at other points $\mathbf{r}'$. Neglecting nonlinear optical effects the dipole moment induced in an infinitesimal volume element $d\mathbf{r}'$ is proportional, through the polarizability α, to the effective field $\mathbf{E}$ and to the number of particles $\rho\, d\mathbf{r}'$ in that volume. (In this sense $\alpha\rho$ corresponds to a "polarizability density.") Now, the electric field at the point $\mathbf{r}$ at time t due to the $\alpha\rho\, d\mathbf{r}'\mathbf{E}$ dipole at $\mathbf{r}'$ is given by[86]

$$\nabla_{\mathbf{r}} \times \nabla_{\mathbf{r}} \times \frac{\alpha\rho[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]\, d\mathbf{r}'\mathbf{E}[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]}{|\mathbf{r} - \mathbf{r}'|}$$

Here we have shown the explicit dependence of the curl operators on the position $\mathbf{r}$ and of the number density and electric vector on the retarded time $t - |\mathbf{r} - \mathbf{r}'|/c$. The effective field, then, satisfies the integral equation

$$\mathbf{E}(\mathbf{r}t) = \mathbf{E}_0(\mathbf{r}t) + \int_{V-\sigma_{\mathbf{r}}} d\mathbf{r}\, \nabla_{\mathbf{r}} \times \nabla_{\mathbf{r}} \times \frac{\alpha\rho[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]\mathbf{E}[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]}{|\mathbf{r} - \mathbf{r}'|} \tag{5.1}$$

Equation (5.1), together with an accompanying relation for the magnetic field, is equivalent to Maxwell's equations for the fluid system in the DID approximation.[86] The integration excludes a small sphere $\sigma_{\mathbf{r}}$ surrounding the atom or molecule at $\mathbf{r}$: if $\mathbf{r}$ lies outside the volume V, of course, $\sigma_{\mathbf{r}} = 0$.

The function $\rho(\mathbf{r}', t) \equiv \rho(\mathbf{r}', t; \mathbf{R}_1, \ldots, \mathbf{R}_N)$ is to be interpreted as the instantaneous number density at $\mathbf{r}'$ and time t' corresponding to the N-particle configuration $\mathbf{R}_1, \ldots, \mathbf{R}_N$. $\rho(\mathbf{r}', t')$ enjoys the well-known properties:

$$\begin{aligned} \langle \rho(\mathbf{r}', t') \rangle &= n^{(1)}(\mathbf{r}', t') \\ \langle \rho(\mathbf{r}, t)\rho(\mathbf{r}', t') \rangle &= n^{(2)}(\mathbf{r}t, \mathbf{r}'t'), \quad \text{etc.} \end{aligned} \tag{5.2}$$

where $n^{(m)}(\mathbf{r}t, \mathbf{r}'t', \ldots, \mathbf{r}^m t^m)$ is the smoothly varying, canonical ensemble averaged, m-particle density function which gives the probability that some particle is at $\mathbf{r}$ at time t and some particle (*not* necessarily a different one) is at $\mathbf{r}'$ at time t', and so on. For $m = 1, 2$, we have for a homogeneous fluid that $n^{(1)}(\mathbf{r}t) = N/V = \bar{\rho}$, while $n^{(2)}(\mathbf{r}t, \mathbf{r}'t') = \bar{\rho} G^{VH}(|\mathbf{r} - \mathbf{r}'|, t - t')$ where G^{VH} is the Van Hove correlation function. Later it will be useful to break up the instantaneous number density into its uniform and fluctuating parts, that is, $\rho \equiv \bar{\rho} + \Delta\rho$. The only other distribution functions that

arise in our second-order theory are those corresponding to $m = 3$ and $m = 4$.

We can solve the integral equation (5.1) for the effective field by an iterative scheme of successive approximations. The Ewald-Oseen extinction theorem[86] suggests that for a zero-order $\mathbf{E}$ we should take (rather than $\mathbf{E}_0$) the local field satisfying

$$\mathbf{E}_L(\mathbf{r}t) = \mathbf{E}_0(\mathbf{r}t) + \int_{V-\sigma_\mathbf{r}} d\mathbf{r}' \nabla_\mathbf{r} \times \nabla_\mathbf{r} \times \frac{\alpha\bar{\rho}\mathbf{E}_L[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]}{|\mathbf{r} - \mathbf{r}'|} \tag{5.3}$$

$\mathbf{E}_L$ is the effective field for the case in which the density fluctuations are neglected. The solution to (5.3) is the well-known local field which propagates in a uniform medium with a velocity c/n [n, the mean index of refraction, being given by the Lorentz-Lorenz relation $(n^2 - 1)/n^2 + 2) = 4\pi\alpha\bar{\rho}/3$] but with the same frequency as the external field $\mathbf{E}_0$. It is important to note that in general[86,87] $\mathbf{E}_L$ (the electric vector of the refracted wave) is not necessarily parallel to $\mathbf{E}_0$, for example, for z-polarized incident light $\mathbf{E}_L$ will have a nonvanishing (although small—see Section V.C) x-component.

Setting $\mathbf{E} = \mathbf{E}_L$ ($\equiv \mathbf{E}_{(0)}$) on the right-hand side of (5.1) we have that

$$\mathbf{E}(\mathbf{r}t) = \mathbf{E}_0(\mathbf{r}t) + \alpha \int_{V-\sigma_\mathbf{r}} d\mathbf{r}\, \nabla_\mathbf{r} \times \nabla_\mathbf{r} \times \frac{\rho[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]\mathbf{E}_L[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]}{|\mathbf{r} - \mathbf{r}'|} \equiv \mathbf{E}_{(1)}(\mathbf{r}t) \tag{5.4}$$

A solution for $\mathbf{E}$ through higher order is then obtained by setting $\mathbf{E} = \mathbf{E}_{(1)}$ on the right-hand side of (5.1):

$$\begin{aligned}\mathbf{E}(\mathbf{r}t) \equiv \mathbf{E}_{(2)}(\mathbf{r}t) &= \mathbf{E}_0(\mathbf{r}t) \\ &+ \alpha \int_{V-\sigma_\mathbf{r}} d\mathbf{r}' \nabla_\mathbf{r} \times \nabla_\mathbf{r} \times \frac{\rho[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]\mathbf{E}_0[\mathbf{r}', t - (|\mathbf{r} - \mathbf{r}'|/c)]}{|\mathbf{r} - \mathbf{r}'|} \\ &+ \alpha^2 \int_{V-\sigma_\mathbf{r}} d\mathbf{r}' \nabla_\mathbf{r} \times \nabla_\mathbf{r} \\ &\quad \rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) \int_{V-\sigma_{\mathbf{r}'}} d\mathbf{r}'' \nabla_{\mathbf{r}'} \times \nabla_{\mathbf{r}'} \rho[\mathbf{r}'', t - (|\mathbf{r} - \mathbf{r}'|/c) \\ &\quad - (|\mathbf{r}' - \mathbf{r}''|/c)] \frac{\mathbf{E}_L[\mathbf{r}'', t - (|\mathbf{r} - \mathbf{r}'|/c) - (|\mathbf{r}' - \mathbf{r}''|/c)]}{|\mathbf{r}' - \mathbf{r}''|} \\ &\times \frac{}{|\mathbf{r} - \mathbf{r}'|}\end{aligned} \tag{5.5}$$

Rather than iterating further we shall show that the effective field given here through second order is sufficient to (*1*) make contact with the

several light scattering theories cited in Section III; and (*2*) calculate the effect of three- and four-particle correlations.

In the above discussion we have neglected the dependence of the polarizability on the interparticle distances. This simplification was employed by Kirkwood,[83] Yvon,[84] Rosenfeld,[85] and others[86] in their statistical mechanical theories of the optical properties of many-body systems. Mazur and co-workers[88] have improved upon this approximation by taking into account the internal structure of the interacting fluid particles. They have shown, in perturbation calculations on helium, that the corrections to the Lorentz-Lorenz and Clausius-Mossotti functions due to this variation in polarizability are of the same order of magnitude as those arising from local field fluctuations. However, writing their result as [see (5.22)]

$$\alpha_{\text{effective}} = \alpha \mathrm{I} + \theta(\alpha^3) \tag{5.6}$$

it is clear that the effects of the internal structure of the fluid particles can be neglected since they do not contribute until third order in our perturbation scheme. This point is considered again in Section V.D.

According to the classical theory[1,2,25] the frequency distribution of the scattered light detected at a point **R** far from the sample volume is given by the Fourier transform of the autocorrelated electric field **E** at **R**. More explicitly, the *i*-polarized contribution to the observed intensity at **R** is given by

$$I_R^i(\omega') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dt e^{i\omega' t}\langle [\mathbf{E}(\mathbf{R}, 0)]_i^* [\mathbf{E}(\mathbf{R}, t)]_i \rangle \tag{I}$$

where the brackets denote an ensemble average over all fluid configurations.

A. First-Order (Polarized) Inelastic Scattering

From (5.4) we have immediately that $\mathbf{E}(\mathbf{R}t)$ is given through first order in the polarizability by ($\mathbf{r} = \mathbf{R} \notin V$ implies that $\sigma_r = 0 = \mathbf{E}_0(\mathbf{r}t)$)

$$\mathbf{E}(\mathbf{R}, t) = \alpha \int_V d\mathbf{r}' \nabla_{\mathbf{R}} \times \nabla_{\mathbf{R}} \times \frac{\rho[\mathbf{r}', t - (|\mathbf{R} - \mathbf{r}'|/c)]\mathbf{E}_0[\mathbf{r}', t - (|\mathbf{R} - \mathbf{r}'|/c)]}{|\mathbf{R} - \mathbf{r}'|} \tag{5.7}$$

As in Section IV we consider the case of *z*-polarized incident light, that is, $\mathbf{E}_0(\mathbf{r}t) = E_0 \hat{z} e^{i\mathbf{k}\cdot\mathbf{r}} e^{-iwt}$ and a scattering geometry in which **k** is parallel to the *x*-axis while **R** lies along the positive *y*-axis. Substituting (5.7) for $\mathbf{E}(\mathbf{R}, t)$ in (I) and using an approximation for the dipole propagator which is valid for $|\mathbf{R}| \gg |\mathbf{r}'|$, all $\mathbf{r}' \in V$, it is easy to show (see Appendix A) that there is no depolarization and that

$$I_{\mathbf{R}}^{zz}(\omega') = \frac{\alpha^2 |E_0|^2}{2\pi R^2}\left(\frac{\omega}{c}\right)^4 NS(K, \Omega) \tag{5.8}$$

where

$$S(K, \Omega) = \int_V d\mathbf{r} \int_{-\infty}^{+\infty} dt e^{-i\mathbf{k}\cdot\mathbf{r}} e^{i\Omega t} G^{VH}(|\mathbf{r}|, t) \tag{5.9}$$

is the familiar double (space-time) Fourier transform of the Van Hove function with $\mathbf{K} = \mathbf{k} - (\omega/c)(\mathbf{R}/R) \equiv \mathbf{k} - \mathbf{k}'$, and $\Omega \equiv \omega - \omega'$. This result, (5.8) and (5.9), agrees with that obtained [compare (4.32)] in our quantum mechanical discussion in Section IV. It is appropriate to emphasize here, however, that this well-known inversion relating the space ($\mathbf{r}$) and time (t) pair correlations and the observed distribution of momentum ($\mathbf{K}$) and energy (Ω) transfers is valid only to the extent that (5.7) approximates the scattered field, that is, to the neglect of the higher-order terms generated by the iterative solution of Maxwell's equations. This simplification is closely related to the assumption of a one-body additive polarizability—see our quantum mechanical discussion in Section IV—and is treated in some detail elsewhere[73] where we consider higher-order cluster expansion corrections to the polarized spectrum. In the case of *neutron* scattering, an equivalent situation arises with the use of the first Born approximation for calculating cross-sections.[74] Although multiple neutron scattering[89] can be discouraged by a judicious choice of experimental set up,[90a] it nevertheless can never be eliminated entirely; the corresponding corrections have been treated most recently by Sköld et al.[90b]

B. Second-Order (Depolarized) Inelastic Scattering

With the right-angle scattering geometry described above, the depolarized spectrum arises from the x-component of the light scattered along the y-axis. Thus we calculate the observed intensity $I_R{}^{XZ}(\omega')$ by evaluating the autocorrelation function of $\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}, t)$). From (5.5) we can write through second order in the polarizability (again $\sigma_r \equiv 0 \equiv \mathbf{E}_0(\mathbf{r}t)$ since $\mathbf{r} = \mathbf{R} \notin V$, and from the above section the linear terms are seen to vanish)

$$\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}t) = \alpha^2 \int_V d\mathbf{r}' \nabla_{\mathbf{R}} \times \nabla_{\mathbf{R}} \hat{x} \cdot \rho[\mathbf{r}', t - (|\mathbf{R} - \mathbf{r}'|/c)] \int_{V-\sigma_{\mathbf{r}'}} d\mathbf{r}'' \nabla_{\mathbf{r}'} \times \nabla_{\mathbf{r}'} \rho[\mathbf{r}'', t - (|\mathbf{R} - \mathbf{r}'|/c) - (|\mathbf{r}' - \mathbf{r}''|/c)] \times \frac{\times \mathbf{E}_0[\mathbf{r}'', t - (|\mathbf{R} - \mathbf{r}'|/c) - (|\mathbf{r}' - \mathbf{r}''|/c)]}{|\mathbf{r}' - \mathbf{r}''|} \times \frac{}{|\mathbf{R} - \mathbf{r}'|} \tag{5.10}$$

Substituting (5.10) for $\hat{x} \cdot \mathbf{E}(\mathbf{R}, t)$ in (I) we find, with the same bind of algebra used in Appendix A,

$$I_R^{XZ}(\omega') = -\left(\frac{1}{2\pi}\right)^3 \frac{\alpha^4}{R^2}\left(\frac{\omega}{c}\right)^2 |E_0|^2 \int_{-\infty}^{+\infty} d\bar{k} \int_{-\infty}^{+\infty} d\bar{k}' \int_V d\mathbf{r}_1 \int_{-\infty}^{+\infty} dt_1$$
$$\times \int_{V-\sigma_{\mathbf{r}_1}} d\mathbf{r}_2 \int_{-\infty}^{+\infty} dt_2 \int_V d\mathbf{r}_3 \int_{-\infty}^{+\infty} dt_3 \int_{V-\sigma_{\mathbf{r}_3}} d\mathbf{r}_4 \int_{-\infty}^{+\infty} dt_4$$
$$\times e^{i(\omega'-\bar{k}c)t_1} e^{-i(\omega-\bar{k}c)t_2} e^{i\bar{k}'ct_3} e^{i(\omega-\bar{k}'c)t_4} \frac{d^2}{dt_3^2} \delta\left(t_3 + \frac{R}{c} - \frac{\mathbf{R}\cdot\mathbf{r}_3}{Rc}\right)$$
$$\times e^{i\omega'(R/c)-(\mathbf{R}\cdot\mathbf{r}_1/Rc)} e^{i\mathbf{k}\cdot\mathbf{r}_2} e^{-i\mathbf{k}\cdot\mathbf{r}_4} [\mathsf{T}_{\bar{k}}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ}$$
$$\times [\mathsf{T}_{\bar{k}'}(\mathbf{r}_3 - \mathbf{r}_4)]_{XZ}^* \langle \rho(\mathbf{r}_1 t_1)\rho(\mathbf{r}_2 t_2)\rho(\mathbf{r}_3 t_3)\rho(\mathbf{r}_4 t_4)\rangle \qquad (5.11)$$

Here $\mathsf{T}_k(\mathbf{r}_1 - \mathbf{r}_2) \cdot \boldsymbol{\mu}_k(\mathbf{r}_2) \equiv (\nabla_{\mathbf{r}_1} \times \nabla_{\mathbf{r}_1} + k^2\mathsf{I})(e^{ik|\mathbf{r}_1-\mathbf{r}_2|}/|\mathbf{r}_1 - \mathbf{r}_2|) \cdot \boldsymbol{\mu}_k(\mathbf{r}_2)$ gives the electric field at $\mathbf{r}_1$ due to the oscillating ($\omega = ck$) dipole $\boldsymbol{\mu}_k$ at $\mathbf{r}_2$. The second-order result (5.11) is, as we shall see shortly, an even simpler expression than it first appears to be. In any case the physical content becomes readily transparent by considering the equation (5.10), or more conveniently its rewritten form (see notation form Appendix A),

$$\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}t) = \alpha^2 \int_V d\mathbf{r}' \int_{-\infty}^{+\infty} dt' \int_{V-\sigma_{\mathbf{r}'}} d\mathbf{r}'' \int_{-\infty}^{+\infty} dt'' \hat{x} \cdot \mathsf{T}(\mathbf{R} - \mathbf{r}', t - t')$$
$$\times \rho(\mathbf{r}'t')\mathsf{T}(\mathbf{r}' - \mathbf{r}'', t' - t'') \cdot \rho(\mathbf{r}''t'')\mathbf{E}_0(\mathbf{r}''t'') \qquad (5.12)$$

for the x-component of the scattered field. Invoking the propagator role of T we see that $\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}t)$ corresponds to radiation that has been (roughly speaking) scattered twice. "First" the electric field $\mathbf{E}_0$ induces a dipole moment along the z-direction in the infinitesimal volume element $d\mathbf{r}''$ at time t''. The z-moment $\alpha\rho(\mathbf{r}''t'')\, d\mathbf{r}''\mathbf{E}_0(\mathbf{r}''t'')$ then sets up a dipolar field [through $\mathsf{T}(\mathbf{r}' - \mathbf{r}'', t' - t'')$] which polarizes the density fluctuation at $\mathbf{r}'$ at time t'. "Finally" the x-component of the latter moment contributes [through $\mathsf{T}(\mathbf{R} - \mathbf{r}', t - t')$] to the depolarized field $\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}t)$ measured at the detector. See, however, the more rigorous discussion in Section VII.C where we make precise the difference between "double" and "collision induced" depolarized light scattering.

We show explicitly in Section V.D—in the case of a dilute to moderate-pressure gas—that the second order scattering theory presented here can be directly related, as is suggested by the above interpretation, to the binary-collision-induced models developed to explain the "translational Raman" spectra of monatomic and spherically symmetric molecular gases. That is, when the four-particle density can be appropriately factored, the second-order theory outlined above reduces to the familiar situation in which light is scattered by the incremental anisotropic polarizability

associated with isolated pairs of interacting atoms or molecules. Similar arguments obtain at increased pressures where the intercollisional interference effects studied by Van Kranendok and Thibeau (see Section VI) and their co-workers can be taken into account by including higher-order correlations in the factored four-particle density. For dense fluids we believe it is necessary to avoid simplifying assumptions about the collision dynamics and local environment and to use instead our general model-independent theory.

Returning to the present discussion we proceed to consider the elastic (static) limit of the second order scattering theory and demonstrate its equivalence to the fluctuating field analyses of Fixman,[56] Mazur,[60] and Bullough.[63]

C. Static Limit: Equivalence to Local Field Analyses

In order to reduce the result, (5.8) and (5.9), to the static limit we simply neglect the time dependence in the Van Hove density, that is, we set

$$G^{VH}(|\mathbf{r}|, t) \to G^{VH}(|\mathbf{r}|, t = 0) = \delta(\mathbf{r}) + \bar{\rho}g(r) \tag{5.13}$$

where $g(|\mathbf{r}|)$ is the dimensionless radial distribution function. Then we can write immediately

$$\int_{-\infty}^{+\infty} d\omega' [I_R^{ZZ}(\omega')]_{\text{static}} = I_R^{ZZ} = \frac{\alpha^2 |E_0|^2}{R^2} \left(\frac{\omega}{C}\right)^4 \bar{\rho} V S(K) \tag{5.14}$$

with ($\mathbf{K} \neq 0$)

$$S(K) = 1 + \bar{\rho} \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} [g(|\mathbf{r}|) - 1] \tag{5.15}$$

identified as the familiar structure factor. Again, the above result agrees exactly with that obtained earlier [compare (4.28) and (4.29)] in the quantum mechanical discussion.

From the above it is clear that an obvious correction to the Einstein turbidity formula [compare (4.33)] comes from improving upon the approximation $S(K) \to k_B T\gamma$, that is, from taking into account the spatial correlations of the fluid particles. Away from the critical point and for optical wavelengths, however, the K-dependent terms in the expansion of the structure factor are negligible. However, there is another correction which, although also small in magnitude, is *qualitatively* important since it gives rise to depolarization. This effect arises from the fact that in deriving the Einstein relation from the molecular theory immediately above we neglected the $I \sim \alpha^4$ contributions. (At this juncture we assume that all third- and higher-order scattering is negligible.)

Including the "doubly scattered" light, the I_R^{XZ} is no longer zero. Instead we have from (5.11), upon dropping the time dependence in the density function $n^{(4)}$ and performing the trivial integrations over t_1, t_2, t_3,

t_4, $\bar{k}$, and $\bar{k}'$:

$$I_R{}^{XZ} \equiv \int_{-\infty}^{+\infty} d\omega' [I_R{}^{XZ}(\omega')]_{\text{static}}$$
$$= \frac{\alpha^4}{R^2}\left(\frac{\omega}{C}\right)^4 |E_0|^2 \int_V d\mathbf{r}_1 \int_{V-\sigma_{\mathbf{r}_1}} d\mathbf{r}_2 \int_V d\mathbf{r}_3 \int_{V-\sigma_{\mathbf{r}_3}} d\mathbf{r}_4$$
$$\times e^{i\mathbf{k}\cdot(\mathbf{r}_2-\mathbf{r}_4)} e^{-i\mathbf{k}'\cdot(\mathbf{r}_1-\mathbf{r}_3)} [\mathsf{T}_{\omega/c}(\mathbf{r}_1-\mathbf{r}_2)]_{XZ} [\mathsf{T}_{\omega/c}(\mathbf{r}_3-\mathbf{r}_4)]_{XZ}{}^*$$
$$\times \{n^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4)\} \tag{5.16}$$

Here $\mathbf{k}' \equiv (\omega'/c)(\mathbf{R}/R)$, the wave (propagation) vector of the elastically scattered light. Now, within the polarization description on which the integral formulation of dielectrics is based, the above theory has been exact. Thus it must be the case that any procedure that accounts for corrections to the Einstein result (without going beyond the linear-dipole approximation) is equivalent to the multiple scattering approach. We consider in particular the anisotropic local field analyses of Fixman and others.

Fixman was concerned with developing a microscopic derivation of the phenomenological theory of light scattering by local fluctuations in the number density. His discussion is based on the earlier work of Kirkwood and Yvon in which the effective field at an arbitrary point in the N-particle system is expressed as an expansion in the polarizability. The linear term is parallel to the incident field and can be shown to give rise to the Einstein expression for the scattered light intensity. The quadratic term involves an anisotropic correction and corresponds to improving upon Lorentz's assumption of a mean polarization associated with the fluid surrounding the point in question. Fixman does not evaluate the contribution of this second-order term to the elastic spectrum. He suggests, however, that (*1*) it is the source of the depolarization of light observed for monatomic and spherically symmetric molecular fluids; and (*2*) since it is small in magnitude compared with the first-order contribution, the consistency of the Einstein phenomenological relation with the microscopic theory is established. We have worked out explicitly the depolarized component of the scattered light intensity arising from Fixman's second-order correction to the Lorentz field. We find (in our notation)

$$I_R{}^{XZ} = \frac{\alpha^4}{R^2}\left(\frac{\omega}{C}\right)^4 |E_0|^2 \int_V d\mathbf{r}_1 \int_{V-\sigma_{\mathbf{r}_1}} d\mathbf{r}_2 \int_V d\mathbf{r}_3 \int_{V-\sigma_{\mathbf{r}_3}} d\mathbf{r}_4$$
$$\times e^{i\mathbf{k}\cdot(\mathbf{r}_2-\mathbf{r}_4)} e^{-i\mathbf{k}'\cdot(\mathbf{r}_1-\mathbf{r}_3)} [\mathsf{T}_{\omega/c}(\mathbf{r}_1-\mathbf{r}_2)]_{XZ} [\mathsf{T}_{\omega/c}(\mathbf{r}_3-\mathbf{r}_4)]_{XZ}{}^*$$
$$\times \{n^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4) - \bar{\rho} g(|\mathbf{r}_3-\mathbf{r}_4|) n^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3)$$
$$- \bar{\rho} g(|\mathbf{r}_1-\mathbf{r}_2|) n^{(3)}(\mathbf{r}_1\mathbf{r}_3\mathbf{r}_4) + \bar{\rho}^2 g(|\mathbf{r}_1-\mathbf{r}_2|) g(|\mathbf{r}_3-\mathbf{r}_4|) n^{(2)}(\mathbf{r}_1\mathbf{r}_3)\} \tag{5.17}$$

The expressions for I_R^{XZ} given by (5.16) and (5.17) are the same except for the appearance in the latter of the extra three terms involving the triplet and pair densities. These contributions arise for the following reason. Recalling the defining equation (5.3) for $\mathbf{E}_L$ we have from (5.5) that $\mathbf{E}_{(2)}$ can be expressed in the alternative form

$$\begin{aligned}\mathbf{E}_{(2)}(\mathbf{r}t) = \mathbf{E}_L(\mathbf{r}t) &+ \alpha\int_{V-\sigma_\mathbf{r}} d\mathbf{r}'\nabla_\mathbf{r}\times\nabla_\mathbf{r} \\ &\times \frac{\Delta\rho[\mathbf{r}', t-(|\mathbf{r}-\mathbf{r}'|/c)]\mathbf{E}_L[\mathbf{r}', t-(|\mathbf{r}-\mathbf{r}'|/c)]}{|\mathbf{r}-\mathbf{r}'|} \\ &+ \alpha^2\int_{V-\sigma_\mathbf{r}} d\mathbf{r}'\nabla_\mathbf{r}\times\nabla_\mathbf{r} \\ &\quad \rho[\mathbf{r}', t-(|\mathbf{r}-\mathbf{r}'|/c)]\int_{V-\sigma_{\mathbf{r}'}} d\mathbf{r}''\nabla_{\mathbf{r}'}\times\nabla_{\mathbf{r}'} \\ &\qquad \times \frac{\Delta\rho[\mathbf{r}'', t-(|\mathbf{r}-\mathbf{r}'|/c)-(|\mathbf{r}'-\mathbf{r}''|/c)]}{|\mathbf{r}'-\mathbf{r}''|} \\ &\qquad\quad \times \frac{\times\,\mathbf{E}_L[\mathbf{r}'', t-(|\mathbf{r}-\mathbf{r}'|/c)-(|\mathbf{r}'-\mathbf{r}''|/c)]}{|\mathbf{r}-\mathbf{r}'|}\end{aligned} \tag{5.18}$$

That is, the vector functions $\mathbf{E}_{(2)}|^{(5\cdot5)}$ and $\mathbf{E}_{(2)}|^{(5\cdot18)}$ are exactly the same provided that $\mathbf{E}_L$ satisfies (5.3): $\mathbf{E}_{(2)}|^{(5\cdot18)}(\mathbf{r}t)$, in the static limit, is essentially the second-order expansion derived in the Fixman-Mazur-Bullough fluctuating field analyses. It is straightforward to show that their macroscopic polarization $\langle\mathbf{P}\rangle$—the average dipole moment per unit volume—obeys an equation identical to (5.3) except for $\mathbf{E}_0$ being replaced by $\alpha\mathbf{E}_0$, that is, $\alpha\mathbf{E}_L$ can be identified as $\langle\mathbf{P}\rangle$. Now, it has been assumed by Fixman (and similarly by Mazur and Bullough) that the mean polarization can be taken as parallel to the incident field. In this case $\mathbf{E}_L$ no longer satisfies (5.3) for general systems of interest (see, for example, Darwin[87] and Born and Wolf[86]) and the vector functions (5.5) and (5.18) will differ by terms of the order α. More explicitly, approximating $\mathbf{E}_L$ by $\hat{z}E_L$ and taking the x-component of $\mathbf{E}_{(2)}(\mathbf{R}t)|^{(5\cdot18)}$, we get only nonzero contributions from the last term—the term in (5.18) with coefficient α vanishes when $\mathbf{E}_{(2)}(\mathbf{r}t)$ is evaluated at the detector. Substituting this expression into (I) and passing to the static limit we find—instead of the single factor $n^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4)$ appearing in (5.16)—the sum $\{n^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4) - \bar{\rho}n^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) - \bar{\rho}n^{(3)}(\mathbf{r}_1\mathbf{r}_3\mathbf{r}_4) + \bar{\rho}^2n^{(2)}(\mathbf{r}_1\mathbf{r}_3)\}$. The terms in the curly brackets involving the triplet and pair densities give the extra $\theta(\alpha^2)$ contributions found in the

Fixman result (5.17), to the extent that the integral

$$\int_{V-\sigma_0} d\mathbf{r} e^{\pm i\mathbf{k}\cdot\mathbf{r}} [\mathsf{T}_{\omega/c}(\mathbf{r})]_{XZ} [g(|\mathbf{r}|) - 1]$$

can be set equal to zero (σ_0 is a sphere excluding an atomic volume at the origin). This integration over $\mathbf{r}$ can be broken up into two parts: (I) $\sigma_0 \leqslant |\mathbf{r}| < 10$ Å; and (II) $|\mathbf{r}| > 10$ Å. For the second interval, $[g(|\mathbf{r}|) - 1] = 0 + \theta(1/N)$ while for the first region $\mathbf{k}\cdot\mathbf{r} \ll 1$ since $|\mathbf{k}| \approx 10^{-3}$ Å^{-1} for optical wavelengths. We can thereby approximate $\exp(\pm_i \mathbf{k}\cdot\mathbf{r})$ by unity and $\mathsf{T}_{\omega/c}(\mathbf{r})$ by the dyadic $(3\hat{r}\hat{r} - \mathsf{I})/r^3$. But then it is trivially seen that the angular integration over $3\hat{r}_x\hat{r}_z[g(r) - 1]/r^3$ gives zero identically. Note that if instead of taking $\mathbf{E}_L(\mathbf{r}) = \hat{z}E_L(\mathbf{r})$ in (5.18) we retain the correct expression through second order [see (5.3)]

$$\mathbf{E}_L(\mathbf{r}) = \hat{z}E_0(\mathbf{r}) + \alpha\bar{\rho}\int_{V-\sigma_\mathbf{r}} d\mathbf{r}' \mathsf{T}_{\omega/c}(\mathbf{r}' - \mathbf{r})\cdot\mathbf{E}_0(\mathbf{r}') \tag{5.18}$$

then the static limit of the autocorrelated $\hat{x}\cdot\mathbf{E}_{(2)}(\mathbf{R}t)|^{(5\cdot18)}$ gives directly the result (5.16). Also the extra three terms in (5.17) can be shown to vanish identically for $\mathsf{T}_{\omega/c}(\mathbf{r}) \to (3\hat{r}\hat{r} - \mathsf{I})/r^3$ consistent with the fact that $\hat{x}\cdot\mathbf{E}_L(\mathbf{r}) = 0$ for systems in which the dipole propagator can be approximated by this short-range $[(\omega/c)r \ll 1]$ form.

Thus our second-order scattering result in the static limit agrees (as it must) with that which follows from an anisotropic local field formulation. One can show similarly that the third-order terms in the expansion of Fixman correspond to triple scattering plus accompanying interference effects, and so on. As mentioned earlier this equivalence derives (virtually by definition) from the general nature of the multiple scattering approach. Nevertheless, as we have noted above, one must be consistent in applying iterative schemes to obtain the effective field by successive approximations: the spurious terms involving the triplet and pair densities can often be of the same order of magnitude as that arising from the $n^{(4)}$ function. It is interesting to note that the $n^{(3)}$ and $n^{(2)}$ terms do not appear in Stephen's analysis even though he has approximated (in our notation) $\mathbf{E}_L$ in (5.18) by $\mathbf{E}_0$ and thereby neglected the $\theta(\alpha)$ x-component of the Lorentz local field. This is because instead of later writing

$$\langle \Delta\rho(1)\,\Delta\rho(2)\,\Delta\rho(3)\,\Delta\rho(4)\rangle = n^{(4)}(1234) - \bar{\rho}n^{(3)}(123) - \cdots$$

he seems to have simply replaced the ensemble average of the product of fluctuations on the left-hand side by the four-particle density function $n^{(4)}(1234)$. The extra scalar factors of $[1 + \theta(\alpha)]$ have been discussed by Bullough and Mazur who treat the perturbation expansion method in the

context of gauge transformations between "in vacuum" and "in medium" multiple scattering descriptions.

Finally, it is interesting to remark, as noted in Section III, paragraph 3, that when the full four-particle density $n^{(4)}$ is expanded in terms of the "distinct" distribution functions $\rho^{(2)}$, $\rho^{(3)}$, and $\rho^{(4)}$ (see Section VI.D) and the latter written as linear combinations of the Ursell cluster factors, (5.16) is seen to be identical to the depolarized intensity derived by Frisch and McKenna[62] (see also Mountain[64] and the discussion in Section VII.C).

D. Relationship to Collision-Induced Pair Scattering

Suppose we start again with (5.10) for the x-component of the scattered field at the detector. Noting the general conditions on the time varying fluctuations discussed in Appendix B and introducing the point-particle form for the single-particle density, (5.12) can be rewritten in the form

$$\hat{x} \cdot \mathbf{E}_{(2)}(\mathbf{R}t)$$
$$= \frac{1}{R}\left(\frac{\omega}{c}\right)^2 \sum_i \sum_{j \neq i} \left\{ \alpha \left[\mathsf{T}_{\omega/c}\left(\mathbf{R}_i\left(t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right) - \mathbf{R}_j\left(t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right)\right) \right.\right.$$
$$\left.\left. \times \alpha \mathbf{E}_0\left(\mathbf{R}_j\left(t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right), t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right) \right]_x \right\} \quad (5.19)$$

The expression in curly brackets is recognized as the x-component of the dipole induced in the particle i [at time $t - (R/c) + (\hat{R} \cdot \mathbf{r}'/c)$] by the z-moment of j arising from the incident field: we label this quantity $\mu_x[ij; t - (R/c) + (\hat{R} \cdot \mathbf{r}'/c)]$. Then, invoking the stationarity property of the autocorrelated field at the detector we can write

$$I_{\mathbf{R}}{}^{XZ}(\omega') = \frac{1}{R^2}\left(\frac{\omega}{c}\right)^4 \int_{-\infty}^{+\infty} dt e^{i\omega' t} \left\langle \sum_i^N \sum_{i' \neq i}^N \sum_{j \neq i}^N \sum_{j' \neq i'}^N \mu_x{}^*(ij; 0)\mu_x(i'j'; t) \right\rangle \quad (5.20)$$

Note that (5.19) is just a rewriting of the second-order result (5.10) where we have performed some straightforward manipulation to get rid of the explicit dependence on retarded times. $I_{\mathbf{R}}{}^{XZ}(\omega')$ given by (5.20) now admits of several simple physical interpretations which show the precise relationships between the above second-order scattering (and fluctuating internal field) theory with the collision-induced descriptions.

Mazur and co-workers[88] have shown that the induced dipole moment of a fluid particle i perturbed by its neighbor j can be expressed in terms of the free-atom polarizability α and incident field $\mathbf{E}_0$ as (in our notation)

$$\boldsymbol{\mu}(i) = \boldsymbol{\alpha}_{\text{eff}}[\mathbf{E}_0(\mathbf{R}_i) + \mathsf{T}_{\omega/c}(\mathbf{R}_i - \mathbf{R}_j) \cdot \boldsymbol{\mu}(j)] \quad (5.21)$$

where

$$\boldsymbol{\alpha}_{\text{eff}} = \alpha[\mathsf{I} + C_1\alpha^2\mathsf{T}_{\omega/c}(\mathbf{R}_i - \mathbf{R}_j) \cdot \mathsf{T}_{\omega/c}(\mathbf{R}_j - \mathbf{R}_i)$$
$$+ C_2\alpha^2\mathsf{T}_{\omega/c}(\mathbf{R}_i - \mathbf{R}_j) : \mathsf{T}_{\omega/c}(\mathbf{R}_j - \mathbf{R}_i)\mathsf{I} + \cdots \quad (5.22)$$

(C_1 and C_2 are perturbation theoretic constants which have been evaluated only in the case of helium.) Then, recalling that $\mathbf{E}_0$ is the same for i and j since optical wavelengths greatly exceed the relevant interparticle distances, we have trivially that the x-component of $\boldsymbol{\mu}(i)$ can be written as

$$\mu_x(i) = \alpha[\mathsf{T}_{\omega/c}(\mathbf{R}_i - \mathbf{R}_j) \cdot \alpha\mathbf{E}_0(\mathbf{R}_j)]_x + \theta(\alpha^3) \tag{5.23}$$

Finally, since the dipole propagator is independent of the sign of its argument, we can put $\mu_x(i) = \mu_x(j)$ and interpret $\mu_x(ij)$ in (5.20)—through terms of $\theta(\alpha^3)$—as half the x-component of the dipole moment induced in the pair ij of particles by the z-polarized incident field.

In the limit of low density for which only two's of particles are correlated, the only terms in (5.20) that contribute to inelastic scattering are those for which $i' = i$ and $j' = j$. Then (dropping proportionality constants except for those involving N)

$$I_{\mathrm{R}}{}^{XZ}(\omega) \sim N^2 \int_{-\infty}^{+\infty} dt e^{i(\omega'-\omega)t} \langle \alpha_{XZ}(0)\alpha_{XZ}(t)\rangle \tag{5.24}$$

where α_{XZ} denotes the XZ-component of the anisotropic polarizability associated with an interacting particle pair (of which there are $\theta(N^2)$).* Equation (5.24) is identical to the binary collision result (4.45) derived in our quantum mechanical discussion and to the XZ-inelastic spectrum which would be calculated for a sample of N^2 anisotropic ($\boldsymbol{\alpha}$) molecules for which spatial correlations can be neglected. (Note, however, that the α_{XZ}'s in the derivation immediately above are constrained to their long-range DID form.) Thus second order scattering by $\theta(N)$ isolated (isotropic) particles is seen to be equivalent to the single scattering by $\theta(N^2)$ collision pairs. The equivalence of these two descriptions of the depolarized intensity has been obscured many times in the literature, most recently where[91] it has been incorrectly asserted that the two "mechanisms" might be discriminated between by looking for different—namely, α^2 versus α^4—dependences on the isolated particle polarizability. Recall, however, from Sections IV and V that $I^{ZX}_{\text{binary collision-induced}} \sim \langle|\alpha_{XZ}(\mathbf{r})|^2\rangle$ and $\alpha_{XZ}(\mathbf{r}) \to 2\alpha^2 T_{XZ}{}^d(\mathbf{r})$ in the DID limit, thereby giving ($I^{XZ} \sim \alpha^4$) agreement with the $I^{ZX}_{\text{second-order local field}}$ theory.

For moderate- to high-pressure gases the terms in (5.20) corresponding to $i' \neq i$ and/or $j' \neq j$ begin to contribute to the observed spectrum. This signals the breakdown of the isolated binary encounter model and the onset of the intercollisional interference effects considered by Van Kranendonk et al.[16,68] and by Thibeau[7] and co-workers. These many-body

* Here, of course, the $e^{-i\omega t}$ factor in the integral arises from the substitution $\mu_x(t) = \alpha_{XZ}(t) E_0{}^z(t)$.

contributions are treated explicitly in Section VI where we return to the full quantum mechanical intensities derived earlier.

VI. ANALYSIS OF DATA: APPRAISAL AND DISCUSSION

In this section we critically review the several calculations and analyses that have been used to interpret the presently available measurements on the depolarized scattering of light by simple fluids.

A. Isolated Binary Collision Model (Gases)

Perhaps the most aggressively pursued development has been the application of the binary collision model to discussions of the inelastic spectrum line shape. As mentioned above, all many-body effects are neglected in this approach and the depolarized scattering by, say, a dilute monatomic gas is attributed wholly to the reradiation of light by the time varying dipole moments induced in (homonuclear) diatomic collision complexes. In this approximation the frequency distribution $I^{ZX}(\Omega)$ is given by [compare (4.45)]

$$I^{ZX}_{(2)}(\Omega) = \frac{I_0}{(R)^2}\left(\frac{\omega}{c}\right)^4 \frac{V}{2\pi}\int_{-\infty}^{+\infty} dt e^{i\Omega t}\langle \alpha_{XZ}[\mathbf{r}(0)]\alpha_{XZ}[\mathbf{r}(t)]\rangle \tag{6.1}$$

where $\mathbf{r}(t)$ denotes the interparticle separation for an arbitrary pair. Taking into explicit account the axial symmetry of the scattering "molecules" we define

$$\beta(r) = \alpha_{\parallel}(r) - \alpha_{\perp}(r) \tag{6.2}$$

where $\alpha_{\parallel}(r)$ and $\alpha_{\perp}(r)$ are the polarizabilities parallel and perpendicular to the symmetry axis. Furthermore let $x = \cos[\theta(t)]$ where $\theta(t)$ is the angle of the molecular axis at time t relative to its value at $t = 0$. Then the depolarized inelastic spectrum can be written in the form

$$I^{XZ}(\Omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dt e^{i\Omega t} C(t) \tag{6.3}$$

where

$$C(t) = \frac{I_0}{(R)^2}\left(\frac{\omega}{c}\right)^4 \frac{V}{60\pi}\langle \beta(0)\beta(t)P_2(x)\rangle \tag{6.4}$$

with $P_2(x) = \frac{1}{2}[3x^2 - 1]$ the second Legendre polynomial. Then by Fourier inversion of (5.3),

$$C(t) = \int_{-\infty}^{+\infty} d\Omega e^{-i\Omega t} I^{XZ}(\Omega) \tag{6.5}$$

and Taylor series expansion of both $C(t)$ and $e^{-i\Omega t}$, we have that

$$\left.\frac{d^n C(t)}{dt^n}\right|_{t=0} = (i)^n \int_{-\infty}^{+\infty} \Omega^n I^{XZ}(\Omega)\, d\Omega \tag{6.6}$$

The right-hand side above is determined directly from the observed line shape, whereas the left-hand side can be calculated from parameterized models for the pair polarizability anisotropy, $\beta(r)$. More explicitly, the first two moments (area and width) of $I^{XZ}(\Omega)$ are given by

$$C(0) = a \int_{-\infty}^{+\infty} [\beta(r)]^2 g(r) r^2\, dr \tag{6.7}$$

and

$$\left.\frac{d^2 C(t)}{dt^2}\right|_{t=0} = a \frac{k_B T}{m} \int_0^{\infty} \left[\left\{\frac{d\beta(r)}{dr}\right\}^2 + 6\left\{\frac{\beta(r)}{r}\right\}^2\right] g(r) r^2\, dr \tag{6.8}$$

where

$$a = \frac{I_0}{(R)^2}\left(\frac{\omega}{c}\right)^4 \frac{1}{15}$$

All odd moments vanish identically since the correlation function $C(t)$ is even in time.

Levine and Birnbaum[14b] have assumed as an empirical approximation that the polarizability anisotropy associated with an interacting pair of spherical particles can be written in the form

$$\beta(r) = 6\alpha^2 r^{-3} + Br^{-p} \tag{6.9}$$

where B and p are unknown parameters characterizing the short-range electronic overlap correction to the dispersion (dipole-induced-dipole) term. The uncertainties in the observed values of the moments derive from several sources, for example: (*a*) relative rather than absolute intensities are measured so that the integrated spectrum must be normalized by the second Kerr virial coefficient—compare Section VII.B and the discussion below in this section; and (*b*) the strong polarized scattering of laser light from the gas cell requires that $I^{XZ}(\Omega)$ be extrapolated (by, say, an exponential function) in order to obtain the line shape for frequency shifts smaller than about 5 cm^{-1}. Levine and Birnbaum have nevertheless attempted to fit the parameters in (6.9) to the experimental results for the first two moments of the depolarized spectrum; for the noble gases they find that each pair $C(0)|_{exp}$ and $C^{(2)}(0)|_{exp}$ leads to *two* sets of values for B and p. In the case of xenon, for example, both B's are negative and of the same order of magnitude, but one p is about 9.5 while the other is as small as 3.5. The latter solution is disregarded on the basis that it is physically unreasonable for an overlap effect, and it is concluded that short-range repulsive interactions are in fact responsible for significant

contributions to the induced anisotropy. This argument, however, is based on somewhat circular reasoning and, as Levine and Birnbaum have themselves cautioned, the experimental uncertainties involved are of critical importance. It is found in fact that there are values for $C(0)|_{\exp}$ and $C^{(2)}(0)|_{\exp}$ within the range deriving from the approximately 15% error in B_K (the second Kerr virial coefficient) which do not lead to *any* solution for the polarizability parameters B and p. Also, Lallemand[18] has shown that $\beta(r)$'s that fit the area and width of the line shape can fail dramatically to give agreement between calculated and observed *higher* moments.

It should be noted that the calculations of Levine and Birnbaum were based exclusively on the Lennard-Jones 12-6 potential, the latter entering through their choice of $g(r) \approx \exp[-u(r)/k_BT]$ for the equilibrium pair function over which the averages in (6.7) and (6.8) are performed. Lallemand[18] has undertaken an extensive analysis in which he considers the possible role of interparticle force expressions in explaining the discrepancy between experimental and (DID) theoretical results for the integrated depolarized intensity. The several potentials which he treats are (*1*) Kihara; (*2*) Barker-Pompe; (*3*) Dymond-Alder; and (*4*) Lennard-Jones n-6. Also, he parameterizes the short-range term in the polarizability anisotropy by $\lambda \exp(-r/d)$ in addition to the inverse power form Br^{-p}; most importantly, however, λ and d, and B and p, are required to fit the fourth and sixth moments as well as the zeroth and second. Lallemand finds for example that for a Barker-Pompe $u(r)$ and Br^{-p} overlap correction there is good agreement between calculated and observed values of $C(0)$ and $C^2(0)$ for argon at $T \approx 20°C$ and $P \approx 60$ atm with *both* ($B = -1.22$, $p = 4.0$) *and* ($B = -1.25$, $p = 10.1$). The second set, however, fails to give reasonable estimates for $C^{(4)}(0)|_{\exp}$ and $C^{(6)}(0)|_{\exp}$ and thus must be rejected. Levine and Birnbaum's argument would instead have dismissed the first pair of values since $r^{-4.0}$ does not properly characterize short-range repulsion. It is perhaps worthwhile to note that the latter study was based on scattering data from a 30 atm gas sample whereas Lallemand's calculations referred to experimental spectra taken at pressures twice as high. Obviously, for a moment analysis to be valid the Fourier inversion (6.3) must provide the correct relationship between the observed line shape and the autocorrelated polarizability anisotropy; otherwise there is no reason at all to expect the parameter fit to be meaningful. It would therefore be very helpful if more reliable depolarization and Kerr data were available for the dilute gas, binary collision regime [$\bar{\rho} \leq 50$ amagats (see Section VI.D below)]. The recent measurements and low-density line shape expansions by McTague, Ellenson, and Hall[8]—see Section VI.D—and by Lallemand[18] comprise an important step in this

direction; it seems fairly certain that the dipole-induced-dipole mechanism, which neglects the negative contribution to XZ-scattering intensities from overlap effects, can not quantitatively describe the observed intensities; this point is pursued in considerable detail in the sections following.

Since the writing of this review, Birnbaum[92a] has treated the binary collision line shape moments on the basis of more recent and accurate noble gas measurements than had been considered in his earlier analyses.[14b] In particular, argon data[6] at a number of pressures were extrapolated to zero density to eliminate the effects of three- and higher-body interactions; that is, the "coefficient" $I^{(2)}(\Omega)$ was determined from a plot of $I^{ZX}|_{\text{exp}} = I^{(2)}(\Omega)\rho^2 - I^{(3)}(\Omega)\rho^3 + \cdots$. The form for the pair anisotropy $\beta(r)$ given by (4.9) is again considered, but now the *fourth* moment is used to discriminate between the several sets of (B, p) values which give the experimental $C^{(0)}(0)$ and $C^{(2)}(0)$. The physically meaningful p is suggested on this basis to be ≈ 4, in agreement with Lallemand (see above). [When, at pressures as low as 60 atm, the moments are fit from the *observed* line shape $I^{ZX}|_{\text{exp}}$ instead of $I^{(2)}$, p is found to be 20% smaller, indicating that many-particle effects are already contributing (see Ref. 21 and Section VI.D).] Again, however, it must be stressed that these fits are very sensitive to the relatively large experimental errors involved: in particular, the lower moments are distorted by the extrapolation of laboratory data to small (<5 cm^{-1}) frequency shifts, while the accuracy of the higher moments is disturbed by the difficulty in measuring the weak intensities in the *far wings* of the depolarized spectrum. Birnbaum and Cohen[92b] treat these ambiguities in some detail. They also interpret preliminary results for the *three* parameter

$$\beta(r) = 6\alpha^2 \left(\frac{A}{r^3} + \frac{B}{r^p}\right)$$

and suggest that the very short-range (e.g., $p \approx 15$) term found in this way corresponds to a real effect that has been masked in the other theoretical analyses.

B. Two-Slope Line Shape

(*B*). It has been suggested that one need not worry so much about the failure of the binary collision description if one concentrates on the large-frequency-shift components in the depolarized line shape. In particular it is argued that the inelastic spectrum should be composed of *two* exponential regions—say, $\Omega < 50$ cm^{-1} and $\Omega > 100$ cm^{-1}—corresponding to the incremental anisotropies due to the dipole-induced-dipole and electronic overlap effects, respectively. Volterra et al.[6,69] have demonstrated these features by recording the XZ-scattering for gaseous argon to frequency

shifts as large as 180 cm^{-1}. At each of several pressures a semilogarithmic plot of the depolarized line shape showed a marked two-slope behavior of the kind mentioned above. When the intensity in each region (at $\Omega =$ 10 cm^{-1} and $\Omega = 105$ cm^{-1}, say) is plotted versus the square of the density one finds that $I_{\text{large }\Omega} \approx \rho^2$ over the range 65 to 225 amagats whereas $I_{\text{small }\Omega}$ grows increasingly more slowly in the same region. These results are interpreted by a two-mechanism theory as follows. The polarizability anisotropy which gives rise to the far-wing scattering is attributed wholly to overlap effects associated with short-range electronic interactions; this "hard" collision distortion is a much stronger function of interparticle distance (namely, a more sharply peaked variation in time) than the dipole-induced-dipole inverse cube effect and thus can account directly for the large-frequency-shift components. Because of its long-range nature the DID mechanism involves increasing cancellation by destructive interference as the density of the gas is raised—this point is discussed in detail below (see Sections VI.D, VI.F, and VI.G). The EO effect, on the other hand, is sufficiently short-range so that even in the many-body interactions characteristic of dense fluids it is unlikely that more than two particles will overlap strongly; the polarizability anisotropy will thus remain "pairwise additive," consistent with the density dependence observed in the far wings. Related discussions of multiple-slope line shape features have been given by Gersten and co-workers.[5,19] Recent molecular dynamics computations (compare Section VI.G) suggest that a "two-mechanism" approach is *not* necessary (although sufficient) to account for this structure.

C. Isolated Binary Collision Model (Liquids)

The persistence of isolated binary collision effects in *liquids* was first suggested by Bucaro and Litovitz,[11] again with specific reference to the large frequency shifts in the depolarized spectrum. They assume that (*a*) the relative motion of a pair of colliding particles is determined wholly by the relevant Lennard-Jones potential and initial velocities, the latter being governed by a gas-kinetic distribution; (*b*) the *short-range* polarizability anisotropy $\beta(r)$ varies as r^{-n} (for monatomic liquids they take $n = 9$ while for spherically symmetric molecular liquids they set $n = 13$ corresponding to the frame distortion being proportional to the repulsive part of the force); and (*c*) due to the close-packed fluid environment only those collisions are considered that have small or vanishing impact parameters. For large frequency shifts the depolarized spectrum is calculated to have the form

$$I^{XZ}(\Omega) \sim \Omega^{2(n-7)/7} \exp\left(\frac{-\Omega}{\Omega_0}\right) \tag{6.10}$$

where Ω_0 is a numerically calculated quantity depending on the temperature, reduced mass, and Lennard-Jones constants. Semilogarithmic plots of $\Omega^{-2(n-7)/7}\, I^{XZ}_{\text{exp}}(\Omega)$ *for large* Ω then provide experimental estimates of Ω_0 for comparison with the theoretical values. Using a 1.5 W Ar^+ laser operating at 5145 Å Bucaro and Litovitz[11] have measured the *XZ*-scattering of light by CCl_4, C_6H_{12}, CH_3OH, C_2H_5OH, H_2O, NH_3, and $CHCl_3$ over a range extending to 500 cm^{-1} frequency shifts; depolarized liquid spectra had been obtained earlier by Gornall et al.[10] for Xe, $C(CH_3)_4$, $Si(CH_3)_4$, $SiCl_4$, $TiCl_4$, $SnCl_4$ and $SnBr_4$ (with a He-Ne laser at 6328 Å) and by McTague et al.[9] for Ar (argon-ion laser at 5145 Å). The agreement between calculated and observed values of Ω_0 was found to be remarkably good in virtually all cases. Most recently Shin[20] has presented the results of the Bucaro-Litovitz description in a more analytical form; the integration over relative velocities is approximated asymptotically by the method of Laplace rather than evaluated numerically. The success of the (EO) isolated binary collision theory in predicting the far wings of the depolarized scattering should be given only qualitative significance, however, because of the basic assumptions involved. As Bucaro and Litovitz have themselves pointed out in the case of *an*isotropic molecular fluids, a spherically symmetric interparticle potential cannot properly describe the interference effects in the orientation—and collision—dependent $\beta(r)$'s. More importantly there is something not altogether consistent about calculating an exponential line shape with, say, a width as small as 18.8 cm^{-1} (namely, argon liquid at 90°K) and then comparing this result with the experimental value of 19.5 cm^{-1} obtained by measuring the slope in the $\Omega > 40$ cm^{-1} region of log $[\Omega^{-4/7} I^{XZ}_{\text{exp}}(\Omega)]$; the same semilogarithmic plot of the four-sevenths power modulated spectrum [$n = 9$ in (4.10)] would give $\Omega_0^{\text{exp}} \approx 11$ cm^{-1} from the derivative at $\Omega \approx 19$ cm^{-1} (recall the multiple-slope effects mentioned earlier). Tabisz, Wall, Shelton, and Ho[13] have most recently focused on the above ambiguities in a systematic treatment of their line shape data for the depolarized collision-induced light scattering by CCl_4 and C_6H_{12}. Specifically, they showed that least-squares fits to (6.10)—*treating n and* Ω_0 *as parameters*—led to quite different results depending upon which frequency region in the spectral wings was studied. In the case of C_6H_{12}, for example, n and Ω_0 were found to assume the values 8, 9, and 13, and 23, 20, and 15, respectively, for the ranges 16 to 100 cm^{-1}, 30 to 80 cm^{-1}, and 40 to 120 cm^{-1}—the Bucaro and Litovitz[11] calculations suggest instead $n \approx 13$ ("frame distortion") and $\Omega_0 \approx 13$. Tabisz et al. also least-squares fit the same date to the more exact (within the isolated binary collision model) expression of Shin's[20]

$$I^{XZ}(\Omega) \sim \Omega^{[2(n-7)/7]+4/19} \exp\{-F\Omega^{12/19} + G\Omega^{6/19}\} \qquad (6.10a)$$

with n fixed at 13 and F and G parameters related to the intermolecular potential, temperature, etc. Again, the values of F and G vary significantly from one part of the tail to another and show large deviations from those predicted theoretically.

The whole concept of using a binary collision model to describe the depolarized light scattering by dense fluids has been challenged most recently and frontally by the machine computations of Berne, Bishop, and Rahman.[70] They used the method of molecular dynamics to "exactly" calculate the autocorrelation function

$$C_1(t) = \left\langle \sum_i \sum_{i'} \sum_{j \neq i} \sum_{j' \neq i'} \beta[r_{ij}(0)]\beta[r_{i'j'}(t)]P_2[\hat{r}_{ij}(0) \cdot \hat{r}_{i'j'}(t)] \right\rangle$$

whose Fourier transform gives the observed line shape [e.g., see (5.20)]. The spherical particles were assumed to interact through a Lennard-Jones 6–12 potential, and two high-density thermodynamic states $S_1 =$ (765 amagat, 88.5°K) and $S_2 =$ (1000 amagat, 300°K) were studied for each of several choices of $\beta(r)$. In addition the functions

$$C_2(t) = \left\langle \sum_i \sum_{i'} \sum_{j \neq i} \sum_{j' \neq i'} \beta[r_{ij}(0)]\beta[r_{i'j'}(t)] \right\rangle$$

and

$$C_3(t) = \left\langle \sum_i \sum_{i'} \sum_{j \neq i} \sum_{j' \neq i'} P_2[\hat{r}_{ij}(0) \cdot \hat{r}_{i'j'}(t)] \right\rangle$$

were similarly calculated—the first depending *only* on the relative *radial* motions of pairs, and the second on the *orientational* (angular) correlations only. For *all* choices of $\beta(r) \sim 1/r^{11}$, $1/r^7$, $1/r^{9.3}$, and $1/r^3 - 0.473(1/r^{9.3})$ and for both S_1 and S_2, $C_3(t)$ was found to closely resemble $C_1(t)$ over the times scale of interest ($0 \leqslant t \leqslant 2 \times 10^{-12}$ sec) whereas $C_2(t)$ dramatically decays more slowly, thereby suggesting that the spectral widths observed for dense fluids are more sensitive to the correlated reorientations of pairs than to the particular radial form of the incremental anisotropy. Furthermore, for all $\beta(r)$ it was found by direct calculation of (6.4) that the isolated binary collision $C(t)$ is in profound *dis*agreement with the "full" [compare $C_1(t)$] autocorrelation function.

It seems reasonable to conclude that the presently available line shape data for the depolarized collision-induced scattering by molecular fluids can not be simply interpreted in terms of an unadulterated isolated binary collision model. From an *operational* point of view, this means that one cannot yet use low pressure measurements to unambiguously test the many possible short-range polarizability anisotropy effects. Also, conceptually, the use of "head on" pair trajectories and gas-kinetic relative velocity distributions can enjoy only a qualitative, phenomenological basis in the case of dense gases and liquids—nevertheless the binary

collision description of electronic overlap effects appears reasonable as a crude model for the shape and density (and temperature) dependence of the *far* wings in low pressure spectra.

Volterra et al.[69] have noted that an extrapolation of the EO tail gives intensities which are far too small to account for the low-frequency part of the depolarized spectra of simple liquids. In the case of CCl_4, for example, it has been suggested that the "excess scattering" be fit to a Lorentzian and attributed to many-body DID effects whose long-range nature can not be described by an isolated pair interaction assumption. This suggestion is, of course, consistent with the usual local field theories in which the dipole-induced-dipole anisotropy at an arbitrary point in the fluid arises from the instantaneous density fluctuations in the surrounding volume. To investigate quantitatively the DID mechanism, however, and specifically, to consider the possibility of probing many-body effects, it is most convenient to return again to the case of low-pressure gases.

D. Virial (Density) Expansions

Putting aside for the moment the problem of line shape, we treat now the question of density-dependent *integrated* intensity. In particular there have been several recent calculations and measurements of the depolarization ratio defined by $\Delta \equiv I^{XZ}/I^{ZZ}$. Δ is found experimentally to increase linearly with density at low gas pressures ($P < 50$ atm for room temperature argon, say) consistent with the fact that for sufficiently dilute systems $I^{XZ} \sim \rho^2$, $\rho \to 0$ (the binary collision limit), and $I^{ZZ} \sim \rho$. For higher pressures the depolarization ratio begins to fall off rather dramatically such that at about 200 atm Δ is already a factor of two smaller than what one would expect from an extrapolation of the low-density behavior. Note that the "corrosion" of the ρ^2 dependence of I^{XZ} is not necessarily inconsistent with the increased importance of *pairwise* electronic overlap due to the fluid becoming more close packed. This latter, short-range effect contributes largely to the *weak* scattering in the *far* wings and, even as it weighs heavily in shaping the properties of the *inelastic* spectrum, competes less effectively with the strong *low*-frequency components in determining the *integrated* intensity.

Recall from (5.16) that the depolarized elastic cross-section I^{XZ} can be written (neglecting EO and nonpairwise-additive contributions to the polarizability anisotropy) as an average of exponentials and dipole propagators over the equilibrium four-particle density defined by (5.2). $n^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4)$ can be rewritten exactly in terms of the more familiar distribution functions $\rho^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) \equiv \rho^n g^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$, $n = 1, 2, 3, 4$. Here, for example, $\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3)$ gives the probability of finding a particle at $\mathbf{r}_1$, a different one at $\mathbf{r}_2$, and still a third at $\mathbf{r}_3$; and similarly for $n = 2$

and 4. The result (5.16) for the integrated depolarized scattering intensity then becomes

$$I^{XZ} = \frac{\alpha^4 E_0^2}{R^2}\left(\frac{\omega}{C}\right)^4\left\{2\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2[\mathsf{T}^d(\mathbf{r}_1-\mathbf{r}_2)]_{XZ}{}^2\rho^{(2)}(\mathbf{r}_1\mathbf{r}_2)\right.$$
$$+4\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\int_V d\mathbf{r}_3[\mathsf{T}^d(\mathbf{r}_1-\mathbf{r}_3)]_{XZ}[\mathsf{T}^d(\mathbf{r}_1-\mathbf{r}_2)]_{XZ}\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3)$$
$$\left.+\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\int_V d\mathbf{r}_3\int_V d\mathbf{r}_4[\mathsf{T}^d(\mathbf{r}_1-\mathbf{r}_2)]_{XZ}[\mathsf{T}^d(\mathbf{r}_3-\mathbf{r}_4)]_{XZ}\rho^{(4)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{r}_4)\right\} \tag{6.11}$$

Here we have approximated the dipole propagator by its short distance $[(\omega_0/c)r \ll 1]$ form, $\mathsf{T}^d(\mathbf{r}) = (3\hat{r}\hat{r} - 1)/r_3$, and replaced all exponentials with optical wave vectors by unity. The first term in (6.11) is recognized to be identical to the elastic depolarized spectrum derived directly from an assumption of isolated pair interactions—see, for example, (4.44) with $[\alpha_{XZ}(\mathbf{r})] \to 2\alpha^2[\mathsf{T}^d(\mathbf{r})]_{XZ}$, that is, with the polarizability anisotropy evaluated in the DID limit. The last two terms in (6.11) include the many-body corrections from three- and four-particle correlations.

Upon neglecting the last two terms in (6.11) the depolarization ratio $\Delta \equiv I^{XZ}/I^{ZZ}$, with I^{ZZ} given by (4.28) or (5.14), can be written as

$$\Delta^{(2)} = \frac{(\text{constant})\rho\left\{\int_0^\infty r^2\, dr r^{-6} g(r)\right\}}{S(K)} \tag{6.12}$$

The density dependence of Δ has been measured experimentally in the case of gaseous argon at room temperature.[7] To investigate the breakdown of the binary collision model it is sufficient to consider the pressure range from 20 to, say, 200 atm. In this P-T domain the Percus Yevich (PY) approximation[93] is known to provide $g(r; P, T)$ to high accuracy. The depolarization ratio (6.12) has been calculated using the radial distribution functions obtained by solving the PY integral equation for several intermediate densities: the resulting $\Delta^{(2)}$ values are found[21] to give (see Fig. 3) dramatically poor agreement with the experimental Δ. This discrepancy persists even when the multiplicative constant in (6.12) is chosen to fit the laboratory data near the origin, that is, for pressures up to about 50 atm; scaling in this way to reduce the low-density slope has been argued to be equivalent to taking into account short-range electronic overlap corrections to the dipole-induced-dipole polarizability anisotropy. More explicitly, replacing r^{-6} in (6.12) by $(r^{-3} + \mathrm{Ar}^{-P})^2$, with A and p chosen to given the observed values in the $\rho \to 0$ limit (recall that "constant" $= 24\pi\alpha^2/5$) still leads to a calculated depolarization ratio which

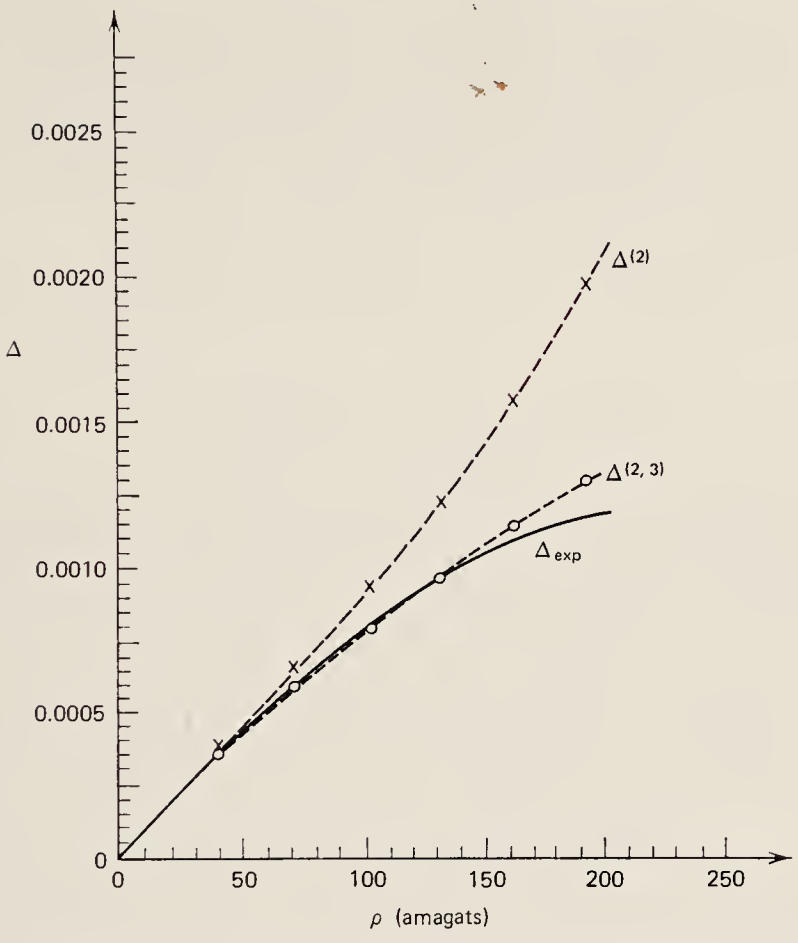


Fig. 3. Depolarization of light by argon gas at room temperature. $\Delta \equiv I^{XZ}/I^{ZZ}$ (see text). $\Delta^{(2)}$ is the theoretical value taking into account only isolated pair interactions, while $\Delta^{(2,3)}$ includes the effects of three-particle correlations as well.

increases with density much faster than do the laboratory measurements. It is therefore reasonable to conclude that the large negative discrepancy between experiment and theory arises simply from the neglect of the many-body corrections. If one assumes further that the magnitude of the three- and four-particle contributions relative to the isolated pair terms can be meaningfully estimated by including only the long-range dispersion effects, then these higher-order correlations can be calculated directly from (6.11).

We consider first the second term in (6.11). In analogy with the Kirkwood superposition approximation $[\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) \approx \rho^3 g(r_{12})g(r_{13})g(r_{23})]$ we write

$$\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) \approx \frac{\rho^3}{3}[g(r_{12})g_0(r_{13})g_0(r_{23}) + g_0(r_{12})g(r_{13})g_0(r_{23}) + g_0(r_{12})g_0(r_{13})g(r_{23})] \quad (6.13)$$

Here $g_0(r)$ is a simple step function which corresponds to the low-density limit of the radial distribution function $g_{HS(\sigma)}(r)$ for hard spheres, with $\sigma = \sigma_{\text{LENNARD-JONES}}$ ($\sigma = 3.4 \times 10^{-8}$ cm for argon). This approximation, the strengths and weaknesses of which have been discussed elsewhere,[21] allows the depolarization ratio $\Delta \equiv I^{XZ}/I^{ZZ}$ to be expressed in the analytic form

$$\Delta^{(2,3)} = \frac{(\text{constant})\rho\left\{\int_0^\infty r^2\, dr r^{-6} g(r) + \frac{4}{\pi}\int_0^\infty dK P^2(K)[S(K) - 1]\right\}}{S(K)} \tag{6.14}$$

where

$$P(K) = K\left[\frac{\sin K\sigma}{(K\sigma)^3} - \frac{\cos K\sigma}{(K\sigma)^2}\right] \tag{6.15}$$

Using PY solutions for the room-temperature radial distribution function $g(r)$, the calculated density dependence of $\Delta^{(2,3)}$ (see Fig. 3) is found[21] to give considerably improved agreement between theory and experiment. The small discrepancy that remains for pressures above 150 atm ($\sim$150 amagats) arises from several sources: (*a*) the contributions from four-particle correlations have not been included and begin to be nonnegligible for these densities; (*b*) the factorization (6.13) provides only an approximate representation of $\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3)$; and (*c*) a dipole-induced-dipole mechanism has been used to describe the polarization of the gas. Nevertheless it appears likely from these results that, for pressures greater than 50 atm, many-body interactions make important negative contributions to the depolarized scattering, corresponding to a lowering of the effective anisotropy in the local fluid environment. This destructive interference would be incorrectly ascribed to pair properties (e.g., short-range polarizability distortion, and interatomic potentials) if a binary collision model were used to interpret much of the presently available data.

Recall that even in the limit of low density where the $\Delta(\rho)$ plot is linear (corresponding to true binary collision dynamics and a density-independent radial distribution function), its slope was found to be more than 30% too large when the polarizability anisotropy was assumed to have the long-range inverse cube form. This result is consistent with the line shape analyses discussed earlier in which the zeroth moment (integrated intensity) of the depolarized spectrum is calculated to be similarly too large unless $\beta(r)$ is allowed to have the form $6\alpha^2 r^{-3} + Br^{-p}$, say. That is, the disagreement between theory and experiment, *at sufficiently low pressures*, can be attributed to the negative ($\beta < 0$) contributions from the short-range ($\rho > 3$) electronic overlap effects. This observation has also been made by Gray and Ralph[17] who consider a virial expansion for the DID depolarization ratio similar to that discussed above. Their coefficient for

the linear term is identical to ours—see (6.12) above where "constant" = $(24\pi\alpha^2/5)$—but they set $S(K) \to 1$ and $g(r) \to \exp[-u(r)/k_BT]$; the theoretical slope is calculated to be 10 and 50% too large for hard-sphere and Lennard-Jones potentials, respectively, the discrepancies again being ascribed to the neglect of repulsive (EO) polarizability distortion. Expanding the radial distribution function according to $g(r) = \exp[-u(r)/k_gT] + \bar{\rho}g_1(r) + \cdots$, the quadratic term in the density expansion of Δ is found to derive primarily from the triplet term in I^{XZ}—see (6.11)—where

$$\rho^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) = \rho^3 g^{(3)}(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) \to \rho^3 e^{-u(r_{12})/RT}\, e^{-u(r_{13})/kT}\, e^{-u(r_{23})/kT}$$

There are also contributions from the density dependence of $g(r)$ in the pair term of I^{XZ} (this effect being proportional to $\int_0^\infty r^2\, drr^{-5}g_1(r)$) and from the density dependence of the structure factor in I^{ZZ} of the denominator (a term involving the product of $\int_0^\infty r^2\, drr^{-6}e^{-u(r)/kT}$ and the second *pressure* virial coefficient B, the latter arising directly from the long-wavelength limit of the $1/(1+x)$, $x \ll 1$ expansion of $1/S(K) \equiv 1/\{1 + \rho\int_V d\mathbf{r}e^{i\mathbf{k}\cdot\mathbf{r}}[e^{-u(r)/kT} - 1]\}$—recall that $B \equiv -\frac{1}{2}\int_V d\mathbf{r}[e^{-u(r)/kT} - 1]$). Gray and Ralph have calculated the ρ^2 term in Δ only for a hard-sphere potential and thus can not make a meaningful *quantitative* comparison with the low-pressure gas data. Even so it is significant that the three-body corrections to the binary collision model are again found to be negative and nonnegligible for pressures greater than about 50 atm.

McTague, Ellenson, and Hall[8] have also considered the virial series for depolarized light scattering by simple gases. As in the quantum mechanical description treated in Section IV above they start with an expression for I^{XZ} which is proportional to the canonical ensemble average over the modulus squared of the XZ component of the electronic polarizability tensor of the *total* system (sample). The latter is developed in powers of the density using a cluster expansion somewhat different from that presented in Section IV: the integrated intensity is written accordingly as (in our notation from Section IV)

$$I^{XZ} = c\left\{\rho^2\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\,|\alpha_{XZ}(\mathbf{r}_1 - \mathbf{r}_2)|^2 \exp\left[\frac{-u(r_{12})}{k_BT}\right]\right.$$

$$+ \rho^3\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\,|\alpha_{XZ}(\mathbf{r}_1 - \mathbf{r}_2)|^2\, g_1(r_{12})$$

$$\left. + \rho^3\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\int_V d\mathbf{r}_3 B(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) \exp\left[-\frac{u(r_{12}) + u(r_{13}) + u(r_{23})}{k_BT}\right]\right\} \tag{6.16}$$

Here

$$B(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) = \alpha_{XZ}(\mathbf{r}_{12})\alpha_{XZ}(\mathbf{r}_{13}) + 2\alpha_{XZ}(\mathbf{r}_{12})\bar{\alpha}^{(3)}_{XZ}(\mathbf{r}_{12}, \mathbf{r}_{13}) + \frac{|\bar{\alpha}^{(3)}_{XZ}(\mathbf{r}_{12}, \mathbf{r}_{13})|^2}{2}$$

and

$$\bar{\alpha}^{(3)}_{XZ}(123) = {}^{\text{irr}}\alpha_{XZ}(123) + \alpha_{XZ}(23) \tag{6.16a}$$

with ${}^{\text{irr}}\alpha_{XZ}(\mathbf{r}_{13}, \mathbf{r}_{12})$, as before, the XZ irreducible triplet component of the fixed-nuclei polarizability associated with a cluster of three atoms having relative positions defined by the vectors $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{r}_{13} = \mathbf{r}_1 - \mathbf{r}_3$. Since it is very difficult to work directly with the nonpairwise additive term (see, however, Section VII.B below) its effect is neglected and the remaining terms in (6.16) are evaluated numerically for a Lennard-Jones potential (argon, $T \approx 300°\text{K}$) within the DID approximation; the ρ^3 terms are again found to be negative (dominated by the three-body interference effects) and nonnegligible—they are estimated in fact to exactly cancel the ρ^2 binary collision effects at a density of 284 amagats (atm). As McTague et al. note, of course, at these pressures the $\theta(\rho^4)$ terms will already be significant. These workers have analyzed the two- and three-particle contributions to the depolarized scattering by a least-squares fit of their experimental data to

$$\frac{I^{XZ}(\Omega)}{\rho^2} = I^{XZ}_{(2)}(\Omega) + \rho I^{XZ}_{(3)}(\Omega) \tag{6.17}$$

$I^{XZ}_{(3)}(\Omega)$ was found to be negative for all frequencies in the range ($7.5\ \text{cm}^{-1} \leq \Omega \leq 50\ \text{cm}^{-1}$) for which it was nonnegligible, and to be characterized by a breadth of the order of $8\ \text{cm}^{-1}$. The observed two-body line shape, on the other hand, extended beyond $150\ \text{cm}^{-1}$ into the wings and, as suggested by Volterra et al. (see above), was well represented by a sum of two exponentials with widths of 12 and $22\ \text{cm}^{-1}$. It has been argued that these room-temperature results imply a three-particle correlation time about 50% longer than that associated with pair motions; a more rigorous categorization of these effects, however, should follow directly from the molecular dynamics computations described below in Section VI.G.

Using the rotational Raman lines of H_2 as an internal standard McTague et al. have obtained experimental estimates of the absolute depolarized scattering intensities by assuming that the observed two- and three-body line shapes—see (6.17)—can be extrapolated smoothly between 7.5 and $0\ \text{cm}^{-1}$. The results for $I^{XZ}_{(2)}$, for example, have been compared in the case of argon gas at room temperature with the values measured by other

workers: McTague et al. find that

$$\left\{\frac{I_{(2)}^{XZ}[\exp]}{I_{(2)}^{XZ}[r^{-3}]}\right\} = 0.62 \pm 0.06$$

(here $I_{(2)}^{XZ}[r^{-3}]$ is the theoretical value of $I_{(2)}^{XZ}$ [compare the first term in (6.16)] with the polarizability calculated in the DID limit) as opposed to 0.67 ± 0.03 found by Thibeau and co-workers,[7] 0.73 ± 0.11 by Buckingham and Dunmur,[67] and 0.81 ± 0.12 by Lallemand.[4] The zeroth, second, and fourth moments of $I_{(2)}^{XZ}(\Omega)$ are all found to be compatible with an electronic overlap correction of the form

$$\beta(r) = 6\alpha^2[r^{-3} - 0.72 \exp(0.80r)] \tag{6.18}$$

which latter is also claimed to provide a good fit for the three-body spectrum $I_{(3)}^{XZ}(\Omega)$. In view of the several uncertainties which remain in the analysis, (6.18) can still be given only qualitative significance—nevertheless the relatively *long*-range nature of this pair anisotropy suggests the important role to be played by many-body interference effect in higher-pressure gases, and in liquids (compare the discussion in Section VI.G).

E. Second-Order Raman Theory

As discussed earlier the many-body theory of McTague, Fleury, and Dupre,[9]—based on Stephen's second-order Raman description[65] of the depolarized scattering of light by dipole polarizable pairs of collective motions, is essentially equivalent to the fluctuating local field analysis described in Section V. Recall that in adapting the general result (5.16) we rewrote the four-particle density $n^{(4)}$ in terms of the more familiar distributions functions $\rho^{(n)}$, $n = 2$, 3, and 4—see (6.11). Then since we were specifically interested in probing the role of many-body effects in low- to moderate-pressure gases it was appropriate to consider only the three-particle corrections to the binary collision dynamics and to approximate the triplet correlations by a Kirkwood-type superposition. McTague et al.,[9] however, address themselves directly to the liquid case and lump together all the two-, three-, and four-body effects by writing

$$n^{(4)}(1234) \to n^{(2)}(12)n^{(2)}(34) + n^{(2)}(13)n^{(2)}(24) + n^{(2)}(14)n^{(2)}(23) \tag{6.19}$$

This crude factorization allows them to express the depolarized line shape as a (dipole) weighted convolution of wave vector-frequency dependent structure factors, $S(K, \Omega)$; the latter functions are determined by the neutron scattering data of Skold and Larson[90] and the well-known sum rules[94]

$$\int_{-\infty}^{+\infty} d\Omega S(K, \Omega) = S(K)$$

and

$$\int_{-\infty}^{+\infty} d\Omega \Omega^2 S(K, \Omega) = \frac{k^2 k_B T}{M} \tag{6.20}$$

with $S(K)$ taken from the appropriate angle-dependent X-ray cross-sections.[95] McTague et al. thereby calculate a $I^{XZ}(\Omega)$ for liquid argon near the triplet point which is roughly exponential in the wings but which has a width (15 cm^{-1}) about 30% smaller than their observed value (21 cm^{-1}). As they suggest the discrepancy is largely due to the simplifying factorization [compare (6.19)] used to approximate the four-particle density; we note also, however, the ambiguities discussed earlier associated with short-range (*non*-DID) interactions and the multiple-slope shape of the liquid spectra.

F. Lattice Occupancy Models

Recall that the three-body DID correlations considered earlier accounted for the discrepancy between theory and experiment in the case of the density dependence of the integrated intensities (or depolarization ratio $\Delta \equiv I^{XZ}/I^{ZZ}$) for low- to moderate-pressure gases. In particular these destructive interference effects were found to make significant negative contributions to the XZ scattering, corresponding to a lowering of the effective anisotropy in the local fluid environment. The sign and qualitative pressure dependence of the three-particle correlations are obvious upon inspection of (6.14) since $P^2(K)$ is sharply peaked at $K \approx \pi/\sigma \approx 10^8$ cm^{-1} and the static structure factors $S(K) - 1$ are known to be negative and to grow in magnitude with increasing density in this wave number region. The relationship between depolarization strengths and configurational asymmetry has been taken most literally by Thibeau and co-workers,[7] who treat phenomenologically the role of many-body effects within the context of the fluctuating internal field theory. Noting that the early Buckingham and Stephen[57] calculations take into account only the isolated pair correlations appropriate to low-density gases, they derive a local field expression which is the $k \to 0$, static limit of (5.20)—the dipole moments being evaluated in the DID approximation, for example,

$$\mu_X(ij) = \alpha^2 E_0 [\mathsf{T}^d(ij)]_{XZ}$$

They proceed to evaluate the contributions from terms with $i' \neq i$ and/or $j' \neq j$ by introducing various lattice occupancy models for the immediate environment of a fluid particle. In the "vacancy" theory,[7] for example, it is assumed that the set of all possible molecular positions form the sites of a lattice, the fixed geometry of which remains the same at all densities. The probability that a site is filled is taken to be ρ/ρ_0 where ρ_0 is a density characteristic of the liquid state; these lattice statistics are then shown to

give a depolarization ratio of the form

$$\Delta_{\text{vacancy}} = \Delta^{(2)}\left(1 - \frac{\rho}{\rho_0}\right)^2 \tag{6.21}$$

with $\Delta^{(2)}$ the pair result derived earlier; see (6.12). The factor $[1 - (\rho/\rho_0)]^2$ represents a negligible correction for low-pressure gases (50 atm, say, or less) while for $\rho \approx \rho_0$, $\Delta_{\text{vacancy}} \approx 0$, corresponding to the fact that for these liquid densities all lattice sites are occupied and the local environment is perfectly symmetric; in the region of intermediate anisotropy the depolarized scattering passes through a maximum. It is found that for $\rho_0 \approx 800$ amagats (6.21) gives good agreement with the experimental data obtained for room-temperature argon over a range of 600 amagats. Whereas this description provides a suggestive qualitative picture of the onset of order with increasing density it suffers at the same time from its use of a purely empirical parameter, ρ_0, which is devoid of precise physical meaning. More explicitly, for low-to-moderate-pressure gases the model is superseded by the microscopic calculations mentioned earlier in this section, whereas for sufficiently dense gases, and liquids, it implies a vanishing depolarization in obvious contradiction with experiment. To probe the relationship between XZ scattering and configurational asymmetry in the latter physical situations requires a more complete accounting of the particle correlations. The computation of the anisotropic density fluctuations has been formulated in Sections IV and V as a wholly statistical mechanical problem; namely, the depolarization was expressed—see, for example, (6.11)—in terms of averages over two-, three-, and four-particle densities. For high-pressure gases, and liquids, it is no longer appropriate to approximate $\rho^{(3)}$ and $\rho^{(4)}$ by crude factorizations. Instead Alder, Weis, and Strauss[22] have shown that the many-body contributions to Δ can be evaluated exactly (numerically) by the method of molecular dynamics.[96] Their calculations have been performed primarily for hard spheres, but a few systems have also been treated in which the atoms interact through square-well and Lennard-Jones potentials as well.

G. Molecular Dynamics "Experiments"

Note from the long-wavelength, static DID limit of (5.20) that the XZ-scattering intensity per unit atom is proportional to the dimensionless quantity

$$S^{XZ} = \frac{\sigma^6}{N}\left\langle \left(\sum_{i \neq j}\sum [\mathsf{T}^d(ij)]_{XZ}\right)^2 \right\rangle \tag{6.22}$$

where, as usual, the brackets denote an equilibrium (canonical ensemble) average for a system of N spheres of diameter σ. It is convenient to break

up (6.22) into its pairwise, triplet, and quadruplet terms, namely, $S^{XZ} = S_2{}^{XZ} + S_3{}^{XZ} + S_4{}^{XZ}$, where $S_2{}^{XZ} = 2(\sigma^6/N)\left\langle \sum\sum_{i\neq j} (T_{ij})^2 \right\rangle$, $S_3{}^{XZ} = 4(\sigma^6/N)\left\langle \sum\sum\sum_{i\neq j\neq k} T_{ij}T_{jk} \right\rangle$ and

$$S_4{}^{XZ} = \frac{\sigma^6}{N}\left\langle \sum\sum_{i\neq j}\sum\sum_{k\neq l} T_{ij}T_{kl} \right\rangle$$

with $\mathsf{T}_{ij} \equiv [\mathsf{T}^d(\mathbf{R}_i - \mathbf{R}_j)]_{XZ}$; these separate contributions are exactly equal (except for trivial multiplicative constants characterizing the scattering experiment) to the successive terms in (6.11). The averages defining $S_2{}^{XZ}$, $S_3{}^{XZ}$, and $S_4{}^{XZ}$ have been evaluated[22] individually for a system of 108 hard spheres at a density corresponding to a reduced volume per atom (V/N) of $1.6\sigma^3/(2)^{1/2}$ and found to be $+7.69$, -13.90, and $+6.39$, respectively, leaving a total of $+0.17$. The near cancellation of pair and many-body effects is consistent with the fact that the measured values of depolarized intensities for liquids are more than an order of magnitude less than that expected by extrapolation from the low-pressure gas (binary collision) regime. S^{XZ}_{total} and $S_2{}^{XZ}$ were also computed at each of several densities intermediate between $V/N = 1.6\sigma^3/(2)^{1/2}[\rho^* \equiv (N/V)\sigma^3 = 0.885]$ and $\rho^* = 0.071$ (71 amagat). In the low-density limit the pair term dominates, giving rise to the observed linear asymptotic behavior for S^{XZ}_{total}. At higher densities S^{XZ}_{total} grows more slowly, reaches a maximum, and then decreases due to the interference between pair, triplet, and quadruplet terms (recall that $S_2{}^{XZ}$increases monotonically). Calculations were also carried out for a greater number of particles (500) and for a square-well potential $(k_BT/\epsilon \approx 1)$ and S^{XZ}_{total} found to be only slightly modified in the region of liquid densities.

Recalling the proportionality constants which relate S^{XZ} [see (6.22)] to the depolarized intensity, that is, writing

$$I^{XZ} = \alpha^4 \frac{|E_0|^2}{R^2}\left(\frac{\omega_0}{C}\right)^4 \frac{N}{\sigma^6} S^{XZ} \tag{6.23}$$

and invoking the earlier long-wavelength result

$$I^{XZ} = \alpha^4 \frac{|E_0|^2}{R^2}\left(\frac{\omega_0}{C}\right)^4 N k_B T \rho \beta_T \tag{6.24}$$

we have that

$$S^{XZ}\Big|_{\text{"exp"}} = \frac{\sigma^6}{\alpha^2}(k_BT\rho\beta_T)\Delta \tag{6.25}$$

Equation (6.25) provides an "experimental" value for S^{XZ} with which the computed quantity given by (6.22) can be compared; $\beta_T \equiv (1/\rho)(\partial\rho/\partial\rho)_T$ is the isothermal compressibility and Δ the observed depolarization ratio. (Alternatively, $S^{XZ}|_{\text{exp}}$ can be determined from absolute intensity measurements if an "internal standard" is used.) At low densities Alder et al. find that the theoretical slope is too large by as much as 50%, a familiar result already discussed earlier. The discrepancy between the calculations (both for hard-sphere *and* Lennard-Jones potentials) and experiment grows with increasing density and is attributed to the enhanced importance of short-range distortion of the atomic charge distributions. These EO effects are argued to lower (raise) the electron density in regions where the local concentration of particles is high (low); in this way the repulsive distortion offsets the effect of the asymmetric configurational fluctuations and the total polarizability anisotropy is an order of magnitude less at liquid densities than it would be if the atoms were not deformable. Thus the molecular dynamics computations are seen to (*1*) mimic the observed density-dependent "shape" of the depolarization ratio (i.e., linear increase near the origin, $\rho \to 0$, and a maximum near the critical density of the fluid); (*2*) suggest the *non*cancellation of long-range DID many-body effects at liquid densities even though it is these same destructive interferences that cause the breakdown of the binary collision model in the moderate- to high-pressure gas regime; and (*3*) point up the importance of short-range EO effects in determining the polarizability anisotropy that gives rise to the XZ scattering of light by simple fluids.

The situation with regard to the depolarized *line shape* has been similarly studied[97] by molecular dynamics computation of the autocorrelation function of the dipolar asymmetry arising from instantaneous density fluctuations. More explicitly, Alder, Strauss, and Weis[97] have evaluated the time dependence of

$$\rho^{XZ}(\tau) = \frac{\left\langle \left[\sum\sum_{i\neq j} T_{ij}{}^{XZ}(0)\right]\left[\sum\sum_{i'\neq j'} T_{i'j'}^{XZ}(\tau)\right]\right\rangle}{\left\langle \left[\sum\sum_{i\neq j} T_{ij}{}^{XZ}(0)\right]\left[\sum\sum_{i'\neq j'} T_{i'j'}^{XZ}(0)\right]\right\rangle} \tag{6.26}$$

where

$$T_{ij}{}^{XZ} = \frac{1}{r_{ij}{}^3}\left(\delta_{XZ} - \frac{3X_{ij}Z_{ij}}{r_{ij}{}^2}\right)$$

with δ_{XZ} the Kronecker delta function and r_{ij} the distance (with X- and Z-components X_{ij}, Z_{ij}) between particles i and j. Within the statistical errors,

$$\rho^{XZ}(\tau) = \rho^{XY}(\tau) = \rho^{YZ}(\tau) = \rho^{YX}(\tau) = \rho^{ZX}(\tau) = \rho^{ZY}(\tau)$$

and, accordingly, the autocorrelation function reported is an average of these six separate components. Note that (6.26) is identical (apart from trivial proportionality constants) to the long-wavelength, DID limit of the integrand appearing in (5.20); similarly, the band shape is obtained by simple Fourier transformation.

Alder et al. have calculated the autocorrelation function $\rho(\tau)$ for atoms interacting through Lennard-Jones, hard-sphere, and square-well potentials, and for widely varying density and temperature conditions. In the case of hard spheres, for example, $I(\Omega) \sim \int_{-\infty}^{+\infty} d\tau e^{i\Omega t}\rho(\tau)$ is found to depend rather sensitively on temperature for fixed density, suggesting that the experimental band shape can not be interpreted too literally in terms of line broadening mechanisms alone. $\rho(\tau)$ is shown further to be dependent on the particular choice of interatomic potential—for example, for argon at $T = 90°\text{K}$ and $\rho \approx \frac{2}{3}$ close packing, the autocorrelation function falls off increasingly more slowly as one passes from Lennard-Jones, to hard-sphere, and finally to square-well interactions, respectively. This effect has been tentatively explained by the increasing viscosities associated with the corresponding many-body systems, and hence the slower relaxation of the density anisotropy. Also it is expected that $\rho(\tau)$ will be sensitive to small changes in the potential, $u(r)$, if the significant range of interatomic distances involved in the (XZ) induced moments is of the order of the distance over which the attractive part of $u(r)$ changes rapidly. In any case, it is interesting that the *band shapes* calculated from these very different autocorrelation functions are *very similar*. That is, when the calculated (semilogarithmic) $I(\Omega)$ curves are shifted (keeping them parallel to their original direction—i.e., the spectra are uniformly scaled) to bring them into near coincidence with the experimental curve (argon, $T = 90°\text{K}$, $\rho = \frac{2}{3}$ close packing), it is very difficult to distinguish between the various interaction potentials used. Obviously, it will be of interest in this regard when more precise *absolute* intensity data become available and when the measurements can be extended over a broader frequency region. [At lower densities and/or higher temperatures, the strong dependence of $\rho(\tau)$ on $u(r)$ is found to disappear.]

A particularly interesting feature displayed by the molecular dynamics calculations is the break in slope found for the semilogarithmic plot of $I(\Omega)$ in agreement with laboratory observations—see the discussion in Section VI.B; furthermore, shifts in temperature are shown to cause *qualitative* changes in the curvature of the band over the experimentally available range of frequencies. These results suggest that one must interpret very cautiously the "two-mechanism" (namely, EO vs. DID) approaches used for line shape analyses in the case of dense gases and liquids. As Alder et al. have themselves stressed, however, the molecular

dynamics calculations are based on a ("long-range") dipole-induced-dipole form for the polarizability anisotropy which completely neglects the ("short-range") electronic distortion contributions mentioned throughout our earlier discussion. This starting point was shown[22] to result in a predicted integrated intensity which is too large (by a whole order of magnitude at higher densities). From the agreement between calculated and measured band *shapes*, on the other hand, it is tempting (but ill-advised!) to conclude that the *non*-DID corrections are indeed two-body-additive *and* that they vary with interatomic distance in essentially the same way as the dispersional component.

It is interesting to cite in this connection the recent SCF calculations of O'Brien, Gutschick, McKoy, and McTague[80] on the He-He diatom polarizability. They solve the Hartree-Fock equation for two interacting helium atoms in the presence of a static, uniform field which is alternately parallel ($\|$) and perpendicular ($\perp$) to the "internuclear axis"; the resulting wave functions are used to compute the induced dipole moment $\mu_{\|,\perp}(E_{\|,\perp})$ and hence $\alpha_{\|,\perp} \equiv \mu_{\|,\perp}(E_{\|,\perp})/E_{\|,\perp}$. The trace of $\boldsymbol{\alpha}$, to whose average the second dielectric virial coefficient B_d—see Section VII—is directly proportional, is found to decrease with decreasing r and to give (for the first time) a theoretical value of B_d agreeing in magnitude *and* sign with experiment. Also, of more significance for present purposes, the results suggest that

$$\beta(r) - \frac{6\alpha^2}{r^3} \approx -e^{-r/r_0}$$

where $r_0 \approx 0.74$ a.u. and, as before, $\beta(r) \equiv \alpha_{\|}(r) - \alpha_{\perp}(r)$. This "overlap" correction is argued to be relatively long-range (scaling approximately to r^{-3} for $3\,\text{Å} \leqslant r \leqslant 7\,\text{Å}$), in contrast to the hypothesis of Bucaro and Litovitz discussed earlier. Further calculations of this kind on other atomic systems, as well as experimental analyses such as that of McTague, Ellenson, and Hall,[8] would be of considerable help in clarifying the precise nature of the pair anisotropy giving rise to XZ collision-induced scattering by dilute atomic gases. We use the *ab initio* results of O'Brien et al. for $\alpha(\mathbf{R}_1, \mathbf{R}_2)$ in our discussion [compare Section VII.B below] of dielectric functions.

VII. RELATED OPTICAL PROCESSES

A. Anisotropic Molecular Fluids and Depolarized Scattering

Suppose we consider now a fluid composed not of atoms or spherically symmetric particles but rather of molecules characterized by anisotropic polarizabilities. Then it is straightforward to show that our discussion of

Section IV.B follows through just as before except that now (assuming rigid structures, i.e., neglecting *intra*molecular vibrations): (*1*) the nuclear wavefunctions appearing in, say, (4.34) are replaced by, for example, $|\gamma_g{}^i\rangle\,|R_g{}^i\rangle$ where $R_g{}^i$ denotes the (ground electronic) initial *rotational* state; (*2*) the $\{\mathbf{R}_g\}$ specify the c.m. positions *as well as* the individual orientation coordinates; (*3*) the adiabatic quantities A_{XZ} and B_{ZX}—compare (4.35) and (4.36)—derive from neglecting the translational *and* rotational energy differences compared with the corresponding electronic origin separations. That is, except for the trivial restrictions removed immediately above, our previous derivation—through (4.37)—allowed implicitly for perfectly general particle structure and dynamics. Thus we have, in the case of *an*isotropic molecular fluids, that the integrated depolarized scattering intensity is still given by (4.37), but with the brackets now denoting canonical ensemble averages which include rotational degrees of freedom as well as translational centers of mass.

Let $\sum_i A^{(1)}_{XZ}(\mathbf{R}_i \equiv [\mathbf{r}_i^{\text{c.m.}}, \boldsymbol{\Omega}_i \equiv \{\theta_i\varphi_i\chi_i\}])$ be the limiting form assumed by A_{XZ} for the particular N-particle configurations for which all particles are isolated:

$$A^{(1)}_{XZ}(\mathbf{R}_i) = \frac{\omega}{2\hbar}\left(\frac{m}{e}\right)^2 e^{i\mathbf{k}\cdot\mathbf{r}^{\text{c.m.}}}[\hbar\omega\alpha^{(1)}_{XZ}(\boldsymbol{\Omega}_i) + 2e^2Q^{(1)}_{XZ}(\boldsymbol{\Omega}_i)] \tag{7.1}$$

where $\alpha^{(1)}_{XZ}$ and $Q^{(1)}_{XZ}$ are, respectively, the molecular XZ polarizability and quadrupole moment; an identical result obtains for $B^{(i)}_{ZX}$ except that the plus (+) sign in the square brackets becomes a minus (−). Analogous expressions can be written down for the pair irreducible components; the cluster expansion, namely,

$$A_{XZ} = \sum_i A^{(1)}_{XZ}(\mathbf{R}_i) + \sum_{j<k}\sum A^{(12)}_{XZ}(\mathbf{R}_j\mathbf{R}_k) + \cdots \tag{7.2}$$

N molecules $\qquad\qquad \dfrac{N(N-1)}{2!}$ distinct pairs

is then substituted into the "new" (4.34). The resulting *singlet* contributions to the integrated depolarized intensity can be written as

$$I^{ZX}\Big|_{\text{"singlet"}} = \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4\Big\{N\langle[\alpha^{(1)}_{XZ}(\boldsymbol{\Omega}_1)]^2\rangle + \frac{N(N-1)}{2}\langle e^{i\mathbf{k}\cdot(\mathbf{r}_1^{\text{c.m.}}-\mathbf{r}_2^{\text{c.m.}})}\alpha^{(1)}_{XZ}(\boldsymbol{\Omega}_1)\alpha^{(1)}_{XZ}(\boldsymbol{\Omega}_2)\rangle\Big\} \tag{7.3}$$

the two terms corresponding to coherent and incoherent processes, that is, to self- and pair-correlations of the individual molecules. Equation (7.3)

is identical to the I^{ZX} result obtained, for example, by Ben-Reuven and Gershon[33] where they neglect the distortion of the anisotropic particles by the (e.g., dipolar) fields of their neighbors. (See Refs. 33 and 30a for the original treatment of the inelastic scattering of light by fluctuations in molecular orientations and positions.) These latter corrections, of course, correspond to collision-induced effects arising from the higher-than-singlet irreducible cluster components in (7.2).

Consider first the pair terms—they have a form identical to (7.1) except that now $\mathbf{r}_i^{\text{c.m.}} \to \mathbf{r}_{jk}^{\text{c.m.}}$ in the exponential and the polarizability $\alpha_{XZ}^{(12)}(\mathbf{r}_j^{\text{c.m.}} - \mathbf{r}_k^{\text{c.m.}}; \boldsymbol{\Omega}_j - \boldsymbol{\Omega}_k)$ depends explicitly on relative center of mass positions and orientations. The resulting contributions to the I^{ZX} cross-section—compare (4,37) with the above-mentioned modifications—assume the form

$$\frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \left\langle \left| \sum_{j<k}\sum e^{i\mathbf{K}\cdot\mathbf{r}_{jk}^{\text{c.m.}}} \alpha_{XZ}^{(\text{pair})}(\mathbf{r}_j^{\text{c.m.}} - \mathbf{r}_k^{\text{c.m.}}; \boldsymbol{\Omega}_j - \boldsymbol{\Omega}_k) \right|^2 \right\rangle /$$

$$= \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \frac{1}{2}\left\{ \rho^2 \iint d\mathbf{R}_1\, d\mathbf{R}_2 [\alpha_{XZ}^{(\text{pair})}(12)]^2 g_{\text{anis}}^{(2)}(12) \right.$$

$$\left. + 2\rho^3 \iiint d\mathbf{R}_1\, d\mathbf{R}_2\, d\mathbf{R}_3 e^{i\mathbf{K}\cdot(\mathbf{r}_1^{\text{c.m.}} - \mathbf{r}_2^{\text{c.m.}})/2} [\alpha_{XZ}^{(\text{pair})}(12)][\alpha_{XZ}^{(\text{pair})}(13)] g_{\text{anis}}^{(3)}(123) \right\} \tag{7.4}$$

Here the subscript "anis" on the reduced distribution functions reminds us that their arguments depend explicitly on the molecular orientations, for example, "1" $\equiv (\mathbf{r}_1^{\text{c.m.}}, \boldsymbol{\Omega}_1)$. The "cross terms," those involving *both* $A_{XZ}^{(1)}$ *and* $A_{XZ}^{(2)}$, also provide nonvanishing contributions to I^{ZX}

$$\frac{I_0}{R^2}\left(\frac{w}{c}\right)^4 \left\langle \left[\sum_i e^{i\mathbf{k}\cdot\mathbf{v}_i^{\text{c.m.}}} \alpha_{KZ}^{(1)}(\mathbf{R}_1) \right] \left[\sum_{j<k}\sum e^{i\mathbf{k}\cdot\mathbf{v}_{jk}^{\text{c.m.}}} \alpha_{XZ}^{(\text{pair})}(jk) \right]^* \right\rangle /$$

$$= \frac{I_0}{R^2}\left(\frac{\omega}{c}\right)^4 \frac{1}{2}\left\{ \rho^2 \iint d\mathbf{R}_1\, d\mathbf{R}_2 \alpha_{XZ}^{(1)}(1) \alpha_{XZ}^{(\text{pair})}(12) g_{\text{anis}}^{(2)}(12) + \cdots \right\} \tag{7.5}$$

Now, suppose we were to take into account only the "coherent" terms in (7.3); then we would calculate an integrated intensity which was considerably larger than the experimentally determined value. It is conceivable that this discrepancy could be attributed to the neglected "incoherent" term—but the magnitude (and sign) of the latter is difficult to estimate since so little is known about the *pair*-angular correlation function involved. More significantly, still, the above derivation points up several new sources of depolarized scattering, namely, the collision-induced effects—see (7.4) and (7.5)—arising from the two-body irreducible components in our cluster expansion of the (adiabatic) XZ amplitude.

We could crudely evaluate the "squared" terms, (7.4), by neglecting the anisotropy ($\boldsymbol{\Omega}$-dependence) in the distribution functions $g^{(2)}_{\text{anis}}$, $g^{(3)}_{\text{anis}}$, and $g^{(4)}_{\text{anis}}$, thereby reducing these contributions to the simpler form considered earlier [see (4.54)] for fluids composed of spherically symmetric particles. Because of the ambiguities (namely, DID vs. EO "mechanisms") associated with the correct form for the exact $\alpha^{(\text{pair})}$ (with accompanying neglect of ${}^{\text{irr}}\alpha^{(\text{triplet})}$, etc.), and approximations associated with the cross terms (7.5), it is certainly unreasonable to expect quantitative accuracy for I^{ZX}_{calc}. A more interesting application of the above arguments, at least for the purposes of the present discussion, involves the appearance of the two mechanisms (namely, molecular reorientation and *relative* translation, and their interference effects) that contribute to the fluctuations in local anisotropy and hence to the depolarized scattering of light. This observation provides a natural correspondence with many recent discussions in the literature concerning the fine structure of I^{ZX} spectral line shapes.

Consider, then, the various "primary variable" theories[38–41] according to which one writes (in arbitrary units)

$$I^{ZX}(\mathbf{K}, \Omega) = \tfrac{1}{2}\int_{-\infty}^{+\infty} dt e^{-i\Omega t}\langle \delta\varepsilon_{XZ}(\mathbf{K}0)\, \delta\varepsilon_{XZ}(\mathbf{K}t)\rangle \tag{7.6}$$

and assumes the XZ-dielectric tensor component to be an additive linear function of the microscopic collective variables $Q^{(j)}$, that is

$$\delta\varepsilon_{XZ} = \left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^{(1)}}\right\}_{Q\equiv 0} Q^{(1)} + \left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^{(2)}}\right\}_{Q\equiv 0} Q^{(2)} + \cdots \tag{7.7}$$

The $\mathbf{K}$ in the above arguments denotes as usual the $\mathbf{K}$th Fourier component of the $\mathbf{r}$-dependent quantity in question; the $Q^{(j)}(\mathbf{K}, t)$ are "primary variables" insofar as they correspond to physical mechanisms which directly $[\{\partial\varepsilon_{XZ}/\partial Q(j)\}_0 \neq 0]$ cause fluctuations in the dielectric tensor. Suppose for the moment we restrict ourselves to the two microscopic collective variables Q^S and Q^F (the meaning of the superscript labels is explained below). Then, from (7.6) and (7.7), we have for the integrated intensity of XZ scattering

$$I^{ZX}(\mathbf{K}) = \left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^S}\right\}_0^2 \langle |Q^S(\mathbf{K})|^2\rangle + \left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^F}\right\}_0^2 \langle |Q^F(\mathbf{K})|^2\rangle + 2\left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^S}\right\}_0\left\{\frac{\partial\varepsilon_{XZ}}{\partial Q^F}\right\}_0 \langle Q^S(\mathbf{K})Q^F(\mathbf{K})\rangle \tag{7.8}$$

Keyes, Kivelson, and McTague,[98] for example, have used (7.8) in the $\mathbf{K}\to 0$ limit to treat the $\mathbf{K}$-independent elastic depolarization cross-sections for liquids composed of anisotropic particles. They take the first

primary variable to be the molecular orientation coordinate

$$Q^S_{XZ}(\mathbf{K}) = \sum_j e^{i\mathbf{K}\cdot\mathbf{r}_j^{\text{c.m.}}} \alpha^{(j)}_{XZ}(\mathbf{\Omega}_j) = \sum_a \sum_j \left(\frac{aa}{XZ}\right)^{(j)} e^{i\mathbf{K}\cdot\mathbf{r}_j^{\text{c.m.}}} \alpha_{aa} \tag{7.9}$$

Here, of course, α_{aa} ($a = 1, 2, 3$) denote the principal axis components of the polarizability tensor in the molecule-fixed system and $(aa \mid XZ)^{(j)}$ labels the transformation (for the jth particle) from the molecular (a) to the laboratory (XZ) frame. The second primary variable is assumed to depend on intermolecular distance (and hence to be "fast") and is associated alternatively with (*1*) the local electric fields arising from the dipole moments induced on neighboring particles; (*2*) polarizability distortion due to short-range molecular collisions or electronic overlaps; or (*3*) a fluctuating stress tensor. Keyes et al. then proceed to consider the local electric field (DID) mechanism which they argue can be described by the primary variable

$$\left\{\frac{\partial\varepsilon}{\partial Q^{F(1)}}\right\}_0 Q^{F(1)} = \alpha^2 \sum_{j<k} \sum \frac{(3\hat{r}_{jk}\hat{r}_{jk} - \mathrm{I})}{r_{jk}^3} \tag{7.10}$$

where $\mathbf{r}_{jk} = r_{jk}\hat{r}_{jk}$ is the jk-intermolecular c.m. vector and α is the scalar (isotropic) part of the particle polarizability [i.e., $\boldsymbol{\alpha} = \alpha\mathrm{I} + \cdots$]. Equation (7.11), when expressed as $\alpha^2 \sum\sum_{j<k} T_{XZ}{}^d(\mathbf{r}_{jk}) = \sum\sum_{j<k} a^{\text{pair(DID)}}_{XZ}(\mathbf{r}_{jk})$ is recognized to be the long-range dipole-induced-dipole limit of our two-body additive XZ-polarizability. Accordingly, the second term in (7.8) is seen to be identical ($\mathbf{K} \to 0$) to our "pair" contributions written out in (7.4)—as we remarked earlier, in fact, these latter terms reduce to the simpler form considered in Section IV [see (4.54)] when the anisotropy in the two-, three-, and four-particle distribution functions is dropped. Also, with the "slow" variable $Q_{XZ}{}^S(\mathbf{K})$ defined by (7.9), the first term in (7.8)—with $\{\partial\varepsilon/\partial Q_{XZ}{}^S\}_0 = 4\pi/V$—corresponds exactly to our "singlet" terms of (7.3) including self- and pair-correlations of the individual particle orientations. Similarly, our terms [namely, (7.5)] involving *both* $A^{(1)}_{XZ}$ and $A^{(12)}_{XZ}$ [compare (7.2) and (4.34)] are directly analogous to the cross terms appearing in the two-primary-variable description, (7.8).

Recall that the singlet (rotation)—Q^{slow}—and *inter*molecular pair (relative translation)—$Q^{\text{fast}(1)}$—effects are characterized by sufficiently widely separated time scales and hence give rise to distinct frequency features, namely, the sharp and broad lines, in the $I^{\text{dep}}(\Omega)$ band shape. Keyes, Kivelson, and McTague,[98] using the generalized hydrodynamics formalism of Mori,[99] have solved the two linear coupled equations for the dynamical variables Q^S and $Q\mathbf{K}^{(1)}$ and obtained the relevant correlation functions in terms of the appropriate static averages. In focusing on the

$\mathbf{K} \to 0$, *integrated* intensity they estimate, for example, $\langle Q^S Q^S \rangle$ in (7.8) by including only the self-reorientation effects [compare the first term in (7.3)] and $\langle Q^F Q^F \rangle$ by calculating only the two-particle contributions [compare the first term in (7.4) or (4.54)]. The cross correlations in (7.8)—see (7.6) in the DID limit—vanish unless the distribution function $g^{(2)}_{\text{anis}}$ is anisotropic, but can be reasonably approximated by its leading term in the "usual" angular dependence expansion.[100] In this way, Keyes et al. find that crude estimates of the three static averages are not inconsistent with the observed data; a more quantitative test of the theory obviously requires a more accurate evaluation of the three equal-time correlation functions in (7.8).

In particular one needs to include: (*1*) the *incoherent* contributions [see the second term in (7.3)] to the reorientation scattering (these corrections have been argued to have a qualitative significance by Ben-Reuven and co-workers;[33,38] (*2*) the three- and four-particle collision-induced effects [see the second and third terms of (7.5)] (as mentioned earlier in Section VI.G for the spherically symmetric case, these latter terms are known to be of the same order of magnitude as the pair averages); and (*3*) the higher-order anisotropic component of $g^{(2)}_{\text{anis}}$ in the "cross" terms. Note also that $Q^{F(1)}$ above has been assumed to correspond to a long-range (DID) local field polarization mechanism: short-range distortion due to electronic overlap collisions, say, $Q^{F(2)}$, can be taken into account by returning to our cluster expansion description of (7.2) and (7.5) with its exact (adiabatic) anisotropies $A^{(jk)}_{XZ}(\mathbf{R}_j \mathbf{R}_k)$. Finally, as Keyes et al. have pointed out, one can take for $Q^{F(3)}$ the pressure tensor under the assumption that molecular reorientation follows the stress in the medium.

Still more significantly, because of recent measurements of *angle*-dependent, *in*elastic depolarized scattering, it is interesting to consider extensions of the preceding arguments to include evaluations of the $\mathbf{K} \not\equiv 0$ behavior of the *two*-time correlation functions $\langle Q_{\mathbf{K}}{}^S(0) Q_{-\mathbf{K}}{}^S(t) \rangle$, $\langle Q_{\mathbf{K}}{}^F(0) Q_{-\mathbf{K}}{}^F(t) \rangle$, and $\langle Q_{\mathbf{K}}{}^S(0) Q_{-\mathbf{K}}{}^F(t) \rangle$. More explicitly, in experiments[35–37,41] on a wide variety of nonassociated liquids such as nitrobenzene, quinoline, and aniline, the VH-spectra (i.e., where the polarization vector of the incident *or* detected radiation is perpendicular to the scattering plane), for example, consist of a broad ($\Delta\nu^{br}_{1/2} \gtrapprox 5\ \text{cm}^{-1}$) background and a narrow ($\Delta\nu^{na}_{1/2} \approx 0.2\ \text{cm}^{-1}$) line with a dip in the center giving rise to a doublet whose splitting varies as $\sin \theta/2$ (where θ is the angle between initial and final propagation directions), for example, $\Delta\nu_{\text{doublet}} \approx 0.03\ \text{cm}^{-1}$, $\theta = 90°$. The early excitement surrounding these freshly observed features derived from the possibility that they were manifesting the existence of traveling shear (transverse) collective modes. Several thermodynamic and statistical mechanical theories[38–40,41] have

been developed, each of which claims to explain many and/or all of the high-resolution details in the laboratory data. Common to virtually all these formulations are (*1*) an initial choice of the (one, two, or three) primary variables (mechanisms) contributing to the dielectric tensor fluctuations; and (*2*) a derivation of the (hydrodynamic, or generalized Langevin-linear transport) equations of motion governing the two-time correlation functions involving these quantities. These approaches (in which the angle-dependent fine structure has been attributed to the various couplings among the molecular reorientation, propagating shear waves, and/or dissipative transverse angular momentum densities, etc.) have been reviewed most recently and incisively by several investigators in a French conference proceedings.[41]

B. Dielectric and Kerr Functions

At various points in Sections IV and VI it was mentioned that the *relative* depolarization cross-sections I^{XZ} are often normalized by the corresponding (namely, the same simple fluid under the same sample conditions) second Kerr virial coefficients B_K. We briefly motivate below the precise relationship between I^{XZ} and B_K; the many-body treatment of density-dependent electric-field-induced birefringence is seen to be largely analogous to our previous description of collision-induced light scattering effects. (An earlier paper by Hellwarth[15] includes an interesting discussion of this situation.) Similarly, we present here a quantum mechanical formulation of the dielectric "constant" which again raises many of the same conceptual and quantitative questions besetting all related optical phenomena, for example, how can one distinguish between the various features (namely, long- vs. short-range polarization components, two- vs. three- and four-particle correlations, etc.) characterizing the cluster susceptibilities and dynamics? Which contribute nonnegligibly to the measured property in question?

The study of electric-field-induced birefringence has been pursued for almost a century since Kerr first observed[101] that the application of a static electric field **E** to a substance leads to an anisotropy in its refractive index n. More explicitly, let $n_{\|}$ and $n_{\perp}$ denote the refractive indices appropriate to the propagation of light parallel and perpendicular to the direction of **E**: the electrooptic Kerr effect simply corresponds, then, to $n_{\|}(\mathbf{E}) \neq n_{\perp}(\mathbf{E})$. The early Langevin-Born theory[102] of this effect addressed itself to systems of *non*interacting *non*spherically symmetric molecules which are directly oriented by the external field; the magnitude and temperature dependence of the birefringence in these cases are used to deduce the independent particle polarizability anisotropies. Buckingham and Pople[103] have included *non*linear polarization effects as well, and have shown in this

way that sufficiently *high* field strength Kerr data allow one to deduce information about the *hyper*polarizabilities of the molecules, that is, about the anisotropy induced by the applied field itself. The theory has been generalized further[104,105] to take into account pairwise interactions, but this description has been restricted by the induced-multipole, molecular-field starting point which has already been discussed critically at length for the case of collision-induced light scattering.

Suppose we proceed instead by considering the well-known Lorentz-Lorenz relation[86] for the refractive index n of an isotropic medium composed of N particles occupying a volume V and each characterized by an effective polarizability α:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{4\pi N}{3V}\alpha \tag{7.11}$$

Let us regard (7.11) as the limit of an exact equation for n by associating $N\alpha$ with the canonical ensemble average of the system's *total* polarizability $\alpha^{\text{total}}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ in the approximation where the latter (adiabatic) quantity can be written in one-body-additive form. In this spirit we can write for the phenomenological generalization of (7.11):

$$\frac{n^2 - 1}{n^2 + 2} = \frac{4\pi}{3V}\langle \alpha^{\text{total}}(\{\mathbf{R}^N\})\rangle \tag{7.12}$$

Anticipating our later preoccupation with virial representations of the Kerr electrooptic coefficient, it is convenient to restrict ourselves to low-density fluids for which $n^2 + 2 \approx 3$ and $n^2 - 1 \approx 2(n - 1)$. [This approximation can be trivially avoided when we extend our theory to still denser gases, and liquids—we simply retain the full quotient $(n^2 - 1)/(n^2 + 2)$ on the left-hand side of (7.12) and evaluate the many-body ensemble average accordingly.] Suppose we now allow the system to be made anisotropic by applying a constant (uniform) external electric field $\mathbf{E}$. Then the refractive indices parallel and perpendicular to $\mathbf{E}$ are given by

$$n_{\parallel(\perp)} \approx 1 + \frac{2\pi}{V}\langle \mathbf{e}^{\parallel(\perp)} \cdot \boldsymbol{\alpha}^{\text{total}} \cdot \mathbf{e}^{\parallel(\perp)}\rangle \tag{7.13}$$

where $\mathbf{e}^{\parallel}$ and $\mathbf{e}^{\perp}$ are unit vectors in directions parallel and perpendicular to $\mathbf{E}$ and the brackets now denote a canonical ensemble average for the many-body system *in the presence of* the external electric field. Note that $\langle \cdots \rangle$ depends, as we see below, on $\mathbf{E}$ both through α^{total} (via nonlinear polarization terms) and through the interaction energy appearing in the Boltzmann weighting factors.

Suppose we choose $\mathbf{E} \parallel \hat{z}$; then it follows immediately that

$$n_{\parallel} - n_{\perp} \approx \frac{2\pi}{V} [\langle \alpha_{ZZ}^{\text{total}} \rangle - \langle \alpha_{XX}^{\text{total}} \rangle] \tag{7.14}$$

Consider first the canonical ensemble (E_Z-field) averaged ZZ-component of the total system (E_Z-field) polarizability. In the one-body-additive approximation we can write

$$\alpha_{ZZ}^{\text{total}} \approx \sum_i \alpha_{ZZ}^{(i)} = \sum_i [\alpha_{ZZ} + \beta_{ZZZ} E_Z + \tfrac{1}{2}\gamma_{ZZZZ} E_Z{}^2 + \cdots]$$

and thus

$$\langle \alpha_{ZZ}^{\text{total}} \rangle \equiv \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_N \alpha_{ZZ}^{\text{total}}(\mathbf{R}_1, \ldots, \mathbf{R}_N; E_Z) P^{(N)}(\mathbf{R}_1, \ldots, \mathbf{R}_N; E_Z)$$

$$\rightarrow \langle \alpha_{ZZ}^{\text{total}}|_{\text{one-body}} \rangle \approx N[\alpha_{ZZ} + \beta_{ZZZ} E_Z + \tfrac{1}{2}\gamma_{ZZZZ} E_Z{}^2] \tag{7.15}$$

A similar result obtains for $\langle \alpha_{XX}^{\text{total}} \rangle$, but with $\alpha_{ZZ} \rightarrow \alpha_{XX}$, $\beta_{ZZ} \rightarrow \beta_{XXZ}$, and $\gamma_{ZZZZ} \rightarrow \gamma_{XXZZ}$. Specializing now to the case of isotropically polarizable, centrosymmetric particles, that is, $\alpha_{ZZ} = \alpha_{XX} \equiv \alpha_0$, $\boldsymbol{\beta} \equiv 0$, and $\gamma_{XXZZ} = \frac{1}{3}\gamma_{ZZZZ} \equiv \frac{1}{3}\gamma$, we arrive at the following expression for the electrooptic Kerr coefficient ($V_m \equiv$ molar volume $\equiv N_0(V/N)$, where $N_0 \equiv$ Avogadro's number):

$${}_mK \equiv \lim_{E_Z \rightarrow 0} \frac{2(n_{\parallel} - n_{\perp})V_m}{27 E_Z{}^2} \approx \frac{4\pi N_0 \gamma}{81} \tag{7.16}$$

in agreement with the result obtained by Buckingham and Pople[103] in their local field formulation of nonlinear polarization effects in samples of noninteracting spherical particles, for example, dilute monatomic gases.

In order to determine the density-*squared* contributions to $(n_{\parallel} - n_{\perp})/E_Z{}^2$ we must consider the two-body additive terms $\sum_{\substack{i>j \\ \text{distinct pairs}}} \sum \alpha_{ZZ}^{(ij)}(\mathbf{R}_i \mathbf{R}_j)$ which were neglected above and which give rise to

$$\langle \alpha_{ZZ}^{\text{total}}|_{\text{two body}} \rangle = \frac{N(N-1)}{2} \langle \alpha_{ZZ}^{(12)}(\mathbf{R}_1 \mathbf{R}_2) \rangle_{P(\{\mathbf{R}^N\}; E_Z)} \tag{7.17}$$

Writing

$$N^2 P^{(2)}(\mathbf{R}_1 \mathbf{R}_2) = \rho^2 e^{-\phi(\mathbf{R}_1 \mathbf{R}_2)/k_B T} + \theta(\rho^3) \tag{7.18}$$

where

$$\phi(\mathbf{R}_1 \mathbf{R}_2) = \phi_{E_z=0}(\mathbf{R}_1 \mathbf{R}_2) - \tfrac{1}{2}[2\alpha_0 + \alpha_{ZZ}^{(12)}(\mathbf{R}_1 \mathbf{R}_2)] E_Z{}^2 + \theta(E_Z{}^3) \tag{7.19}$$

and assuming that the second term immediately above is small compared with $k_B T$, we have that

$$\langle \alpha_{ZZ}^{\text{total}}|_{\text{two-body}} \rangle \approx \frac{\rho^2}{2} \langle \alpha_{ZZ}^{(12)}(\mathbf{R}_1 \mathbf{R}_2) \rangle^{(0)}$$

$$- \frac{\rho^2 E_Z{}^2}{4 k_B T} \langle [\alpha_{ZZ}^{(12)}(\mathbf{R}_1 \mathbf{R}_2)]^2 \rangle^{(0)} + \theta(\rho^3) + \theta(E_Z{}^3) \tag{7.20}$$

Here the (0) superscript denotes an $E_Z \equiv 0$ canonical ensemble average. Expressing $-\langle \alpha_{XX}^{\text{total}}|_{\text{two-body}} \rangle$ similarly, and writing $a_{ZZ}^{(12)}(\mathbf{r}) = \alpha_{\perp}^{(12)}(r) + \beta(r)\cos^2\theta$ and $\alpha_{XX}^{(12)}(\mathbf{r}) = \alpha_{\perp}^{(12)}(r) + \beta(r)\sin^2\theta\cos^2\varphi$ [where θ and φ are the polar and azimuthal angles of $\mathbf{r}$ and $\alpha_{\perp}(r)$ and $\beta(r)$ are defined as in Section VI], we can perform straightforwardly the angular distributions in the statistical mechanical averages. Upon simplifying we find that $\langle(\alpha_{ZZ}^{\text{total}} - \alpha_{XX}^{\text{total}})_{\text{two-body}}\rangle$ reduces to

$$-\frac{2\pi}{15}\frac{\rho^2}{kT}E_Z{}^2V\int_0^\infty dr r^2\beta^2(r)g_0(r) + \theta(\rho^3)$$

where $\beta(r) = \alpha_{\parallel}(r) - \alpha_{\perp}(r)$ and $g_0(r)$ is the zero-density limit of the (equilibrium) radial distribution function. Thus, combining this result with the one-body-additive contribution derived above, we have that

$$\begin{aligned} {}_mK &\approx \lim_{E_z\to 0}\frac{2(n_{\parallel} - n_{\perp})_{\substack{\text{one- and}\\ \text{two-body}}} V_m}{27E_Z{}^2} \\ &= \frac{4\pi N_0\gamma}{81} - \frac{8\pi^2 N_0}{405kT}\langle\beta^2(r)\rangle^{(0)}\rho \equiv A_K + B_K\rho \end{aligned} \tag{7.21}$$

Consider again our earlier result [(6.3) and (6.4)] giving the line shape for depolarized scattering by a simple fluid in the isolated binary collision (IBC) limit. Now, for collisions in which the "molecular axis" doesn't change orientation while the atoms are interacting [that is, $\theta(t) \approx 0$ for all times t such that $\beta(t) \neq 0$], $P_2(x) \to 1$ and we can write for the *integrated XZ* intensity:

$$I^{XZ}|_{IBC} = \frac{I_0}{(R)^2}\left(\frac{\omega}{c}\right)^4\frac{N^2}{15}\langle\beta^2(r)\rangle \tag{7.22}$$

Comparing (7.22) with (7.21) we have directly that

$$I^{XZ}|_{IBC} = \frac{27V(\omega/c)^4 I_0 k_B T\rho^2}{(R)^2 16\pi N_0}B_K \tag{7.23}$$

Equation (7.23), then, describes the relationship between the isolated binary collision contributions ($I^{XZ}|_{IBC}$) to the depolarized scattering and the independent pair interaction quantity, B_K—the second Kerr virial coefficient. Now, to obtain accurate integrated cross-sections, the line shape must be measured to small frequency shifts ($\Omega \approx 5\ \text{cm}^{-1}$) and extrapolated carefully to zero. Consider, for example, the most recent data and analysis provided by McTague, Ellenson, and Hall[8]—they obtain

absolute values [see the review of this work in Section VI.D] of $I^{XZ}(\Omega)$ for room-temperature argon at each of several Ω in the range 7 to 150 cm^{-1} and for densities varying successively from 20 to 150 amagats. These data are least-squares fit to

$$\frac{I^{XZ}_{\text{exp}}(\Omega)}{\rho^2} = I_2(\Omega) + I_3(\Omega)\rho \tag{7.24}$$

By assuming that $I_2(\Omega) = I^{XZ}_{\text{exp}}(\Omega)|^{\rho\to 0}_{IBC}$ can be smoothly extrapolated over $0 \leq \Omega \leq 7\ cm^{-1}$, estimates are inferred for the absolute *integrated* intensities. The two-(and three-)body spectra derived in this way are not sensitive to the linear form (7.24). At $\Omega = 12.5\ cm^{-1}$, for example, I_2 (and I_3) change by less than 1% (10%) if the least-squares fit of $I^{XZ}_{\text{exp}}/\rho^2$ is made to a quadratic function $[I_2 + I_3\rho + I_4\rho^2]$ instead. More significantly, the error associated with extrapolating the (low-density) exponential line shape to the origin is probably of the order of 15%.[8] At the same time, if one fits the experimentally determined[67] density dependence of $2\Delta n V_m/27E_Z^2$ to $A_K + B_K\rho + C_K\rho^2$ instead of to the linear function, the inferred second virial coefficient for room-temperature argon is lowered by roughly 15%, somewhat improving the agreement with calculated DID estimates.

The coefficient of the ρ-term in (7.21) is identical *in form* to that derived by other workers for the molar Kerr constant. In these latter treatments,[104,105] however, an effective field, induced multipole formulation was used and hence $\beta(r)$ was implicitly restricted to the long-range (DID) form $6\alpha_0^2/r_3$; in (7.21), on the other hand, the pair anisotropy corresponds instead to the *exact* two-body polarizability properties, including all short-range and exchange effects, etc. Recent experimental determinations[67] of B_K for gaseous Ar, Kr, Xe, and SF_6 exhibit significant discrepancies (involving both magnitude *and sign*) with the theoretical values calculated using the DID $\beta(r)$, and attribute them to the inadequacy of the long-range polarization description. In analogy with the collision-induced depolarization, it was suggested that one should give up the effective field concept and compute the molar Kerr constant instead in terms of that for a gas of (axially symmetric) anisotropically polarizable $[\beta_{\text{exact}}(r)]$ molecules [corresponding to the (homonuclear) diatom interacting pairs]; but then, as before, one can not self-consistently account for the higher-order dynamical and electromagnetic interactions which, as we have already argued in the above depolarization examples, can significantly "distort" the optical properties of gases at surprisingly low pressures.

The situation is very similar in the case of dielectric function measurements[106] on, say, monatomic gases and liquids. Here we again write our

"generalized" Clausius-Mosotti relation in the form

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi N}{3V}\alpha \to \frac{4\pi}{3V}\{\langle \alpha^{\text{total}}(\{\mathbf{R}^N\})\rangle\}$$
$$\approx \frac{4\pi}{3V}\left\{N\alpha_0 + \frac{N^2}{2}\langle \alpha^{(12)}(\mathbf{R}_1\mathbf{R}_2)\rangle + \frac{N^3}{6}\langle {}^{\text{irr}}\alpha\,(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)\rangle + \cdots\right\} \quad (7.25)$$

where, of course, $\varepsilon = n^2$ is the dielectric "constant" and α_0, $\alpha^{(12)}$, and ${}^{\text{irr}}\alpha$ denote as before the one-, two-, and three-body additive components of the total polarizability. It is then straightforward to show that the Clausius-Mosotti function (for simplicity we neglect here the wavelength dependence of the refractive index) can be written as

$$\frac{(\varepsilon - 1)/(\varepsilon + 2)}{\rho} \approx A_d + B_d\rho + C_d\rho^2 + \cdots \quad (7.26)$$

where

$$A_d = \frac{4\pi}{3}\alpha_0 \quad (7.27a)$$

$$B_d = \frac{8\pi^2}{9}\int_0^\infty dr r^2[\alpha_{\parallel}^{(12)}(r) + 2\alpha_{\perp}^{(12)}(r)]g_0(r) = \frac{2\pi}{9}\langle \text{Tr}\,\{\alpha^{(12}(r)\}\rangle^{(0)} \quad (7.27b)$$

and

$$C_d = \frac{8\pi^2}{9}\left[\int_0^\infty dr r^2\,\text{Tr}\,\{\alpha^{(12)}(r)\}g_1(r)\right.$$
$$\left. \times \frac{1}{4\pi V}\int d\mathbf{R}_1\int d\mathbf{R}_2\int d\mathbf{R}_3\,{}^{\text{irr}}\alpha(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)g_0^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)\right] \quad (7.27c)$$

Here $g_1(r)$ is defined by the density expansion of the radial distribution function: $g(r) = [g_0(r) + g_1(r)\rho + \cdots]$. $g_1(r_{12})$ can be rewritten exactly in the form $g_0(r_{12})\int d\mathbf{r}_3 f(r_{13})f(r_{23})$ where the f's are the usual Mayer functions.[107]

Of course, $g_0(r) = \exp[-u(r)/kT]$ and

$$g_0^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) = \exp\left[-\frac{u(R_{12})}{kT}\right]\exp\left[-\frac{u(R_{13})}{kT}\right]\exp\left[-\frac{u(R_{23})}{kT}\right] \quad (7.28)$$

is similarly the zero pressure limit of the *triplet* correlation function. Again, $\alpha^{(12)}$ and ${}^{\text{irr}}\alpha$ are the two- and three-body irreducible (incremental) components of the ZZ-polarizability of the overall sample:

$$\alpha_{ZZ}^{(12)}(\mathbf{R}_1\mathbf{R}_2) \equiv \alpha_{ZZ}(\mathbf{R}_1\mathbf{R}_2)\big|_{\text{only particles 1 and 2 present}} - 2\alpha_{ZZ} \quad (7.29)$$

$${}^{\text{irr}}\alpha_{ZZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) \equiv \alpha_{ZZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)\big|_{\text{only particles 1,2, and 3 present}} - \alpha_{ZZ}^{(2)}(\mathbf{R}_1\mathbf{R}_2)$$
$$- \alpha_{ZZ}^{(12)}(\mathbf{R}_1\mathbf{R}_3) - \alpha_{ZZ}^{(12)}(\mathbf{R}_2\mathbf{R}_3) - 3\alpha_{ZZ} \quad (7.30)$$

[α_{ZZ} is, of course, the ZZ-component of an isolated particle's polarizability, that is, $\alpha_{ZZ} \equiv \alpha_{ZZ}(\mathbf{R}_1)\big|_{\text{only particle 1 present}} \equiv \alpha_0$]. Now, the two-body $\alpha_{ZZ}^{(12)}(\mathbf{r})$—involving only the single relative position $\mathbf{r}$—can be treated classically or quantum mechanically in an entirely straightforward way and has received considerable attention in the literature, as has been discussed throughout this chapter. The *triplet* irreducible component ${}^{\text{irr}}\alpha_{ZZ}$, on the other hand, (*a*) depends adiabatically (in the sense of the Born-Oppenheimer separation) on *three* relative positions, and (*b*) is, in addition, a relatively *small* quantity deriving from the difference between much larger and almost equal quantities [compare (7.30)]. Recent attempts[108] to perform ab initio calculations of ${}^{\text{irr}}\alpha_{ZZ}$ for interacting helium atoms have been frustrated by (*a*) and (*b*) above. Nevertheless, as discussed previously (compare Section IV, e.g.) this three-body susceptibility is instrumental for the proper analysis of a host of experimental data on the collision-induced optical properties of simple fluids. In view of the practical difficulties cited above, and, more significantly, for conceptual reasons that become obvious below, we consider the triplet polarizability according to the superposition approximation

$$ {}^{\text{irr}}\alpha_{ZZ}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3) = \frac{1}{\alpha}\left[\alpha_{ZZ}^{(12)}(12)\alpha_{ZZ}^{(12)}(13) + \alpha_{ZZ}^{(12)}(12)\alpha_{ZZ}^{(12)}(23) + \alpha_{ZZ}^{(12)}(13)\alpha_{ZZ}^{(12)}(23)\right] \quad (7.31) $$

Substituting (7.31) into the second term of (7.27c) we obtain the following expression for the third dielectric virial coefficient, involving only the two-body incremental polarizability:

$$ C_d = \frac{8\pi^2}{3}\int_0^\infty r^2\, dr\left[\frac{\alpha_\parallel^{(2)}(r) + 2\alpha_\perp^{(2)}(r)}{3}\right] g_1(r) + \frac{2\pi}{3\alpha}\int d\mathbf{R}_{12}\int d\mathbf{R}_{23}\alpha_{ZZ}^{(2)}(\mathbf{R}_{12})\alpha_{ZZ}^{(2)}(\mathbf{R}_{23})e^{-\phi(R_{12})} \times e^{-\phi(R_{23})}f(R_{13}) + \frac{3}{2\pi\alpha}(B_d)^2 \quad (7.32) $$

In writing (7.32) we have used the fact that $g_0^{(3)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)$ is symmetric in all three pairs of particles so that each term in (7.31) gives an identical contribution to the thermal average. The third term in (7.32) arises from our having written $e^{-\phi(R_{13})}$ as $f(R_{13}) + 1$, with $\phi(r) \equiv u(r)/kT$. Again, $\alpha_\parallel^{(2)}$ and $\alpha_\perp^{(2)}$ are the parallel and perpendicular components of the diatom incremental polarizability, referenced to the "molecule"-fixed frame, that is, $\alpha_{ZZ}^{(2)}(\mathbf{r}) \equiv \alpha_\parallel^{(2)}(r) + \beta^{(2)}(r)\cos^2\theta$ where $\beta^{(2)}(r) \equiv \alpha_\parallel^{(2)}(r) - \alpha_\perp^{(2)}(r)$

and θ is angle between the "internuclear" vector $\mathbf{r}$ and the space-fixed z-axis. (Note that for convenience we have dropped the pair superscript "12," replacing it by the more concise "2.")

The six-dimensional integration in (7.32) can be greatly simplified by introducing a new set of variables; the details of this coordinate transformation are given elsewhere[109] and we quote only the results here. In particular it is found that the integral in question can be rewritten exactly in the reduced form

$$I = +8\pi^2 \int_0^\infty dr r e^{-\phi(r)} \int_0^\infty dR R e^{-\phi(R)} \int_{|r-R|}^{r+R} ds s f(s) g(r, R, s) \quad (7.33)$$

where

$$g(rRs) = \left[\alpha_\perp^{(2)}(r)\alpha_\perp^{(2)}(R) + \frac{\{\alpha_\perp^{(2)}(r)\beta^{(2)}(R) + \alpha_\perp^{(2)}(R)\beta^{(2)}(r)\}}{3} + \frac{\beta^{(2)}(r)\beta^{(2)}(R)}{15}(2\cos^2\theta + 1) \right] \quad (7.33a)$$

with

$$\cos\theta = \frac{r^2 + R^2 - s^2}{2rR}$$

It is important to note that (7.32), with the integral given by the three-dimensional I above, provides an exact expression for the third dielectric virial coefficient [since $\alpha_\perp^{(2)}$ and $\beta^{(2)}$ correspond to the *exact* (see below) components of the two-body incremental polarizability] to the extent that the triplet quantity ${}^{\text{irr}}\alpha$ can be written in superposed form [see (7.31)].

Suppose we now consider the (DID) limit where $\alpha_\perp^{(2)}(r)$ and $\beta^{(2)}(r) \equiv \alpha_\parallel^{(2)}(r) - \alpha_\perp^{(2)}(r)$ are replaced by their long-range (dispersional) forms. As is well-known the parallel and perpendicular components of the diatom incremental polarizability have the asymptotic behavior[110]

$$\alpha_\parallel^{(2)}(r) \to \frac{4\alpha^2}{r^3}\left[1 + \frac{2\alpha}{r^3} + \mathcal{O}\left(\frac{1}{r^6}\right)\right]$$

and (7.34)

$$\alpha_\perp^{(2)}(r) \to -\frac{2\alpha^2}{r^3}\left[1 - \frac{\alpha}{r^3} + \mathcal{O}\left(\frac{1}{r^6}\right)\right]$$

Introducing these results into the relations (7.27b) and (7.29), and (7.32) and (7.33), defining respectively the second and third dielectric virial coefficients, we find after some straightforward algebra that

$$B_d \to \frac{32\pi^2\alpha^3}{3} \int_0^\infty r^{-4} \exp[-\phi(r)]\, dr \equiv B_d^{\text{DID}} \quad (7.35)$$

and

$$C_d \to \frac{32\pi^2\alpha^3}{3}\int_0^\infty dr g_1(r) r^{-4} + \tfrac{128}{15}\pi^3\alpha^3 \int_0^\infty r\,dr \int_0^\infty R\,dR \int_{|r-R|}^{r+R} s\,ds \times e^{-\phi(r)} e^{-\phi(R)} f(s) \frac{[3\cos^2\theta - 1]}{r^3 R^3} + \frac{3}{2\pi\alpha}(B_d^{\mathrm{DID}})^2 \quad (7.36)$$

It is interesting to compare (7.36) with the result obtained in recent formulations of the Kirkwood, "translational fluctuation," theory of the Clausius-Mosotti function. Levine and McQuarrie[111b,c] find, for example,

$$C_d \mid_{LM} = \frac{8\pi}{3}\frac{\alpha^3}{V}\int d\mathbf{r}_1 \int d\mathbf{r}_2 \int d\mathbf{r}_3 \frac{e^{-\phi(r_{12})}}{r_{12}^6} f(r_{13}) f(r_{23}) + \tfrac{160}{15}\pi^3\alpha^3 \int_0^\infty r\,dr \int_0^\infty R\,dR \int_{|r-R|}^{r+R} s\,ds \times e^{-\phi(r)} e^{-\phi(R)} f(s) \frac{[3\cos^2\theta - 1]}{r^3 R^3} + \sigma(\alpha^4) \quad (7.37)$$

Now, the first terms in (7.36) and (7.37) are seen to be identical as soon as one recalls that our $g_1(r_{12})$ can be written as

$$g_1(r_{12}) = e^{-\phi(r_{12})} \int d\mathbf{r}_3 f(r_{13}) f(r_{23}) \quad (7.38)$$

The second terms differ only by a factor of $\frac{5}{4}$ (somebody made an algebraic mistake). Finally, our third and last contribution—which is seen by inspection to be of order α^5 (as opposed to the first two, α^3, terms)—has simply been obscured in the "$\mathcal{O}(\alpha^4)$" of the Kirkwood theory. This third term is indeed negligible in the DID case, as we see below. In general, however, when corrections due to short-range polarizability distortion are taken into account, this is no longer true and so we must be careful to retain this term.

Note that we have recovered the classical effective field result for the dielectric function of a simple fluid only after (*a*) superposing the irreducible triplet polarizability according to (7.31), and (*b*) replacing the exact two-body components by their long-range form. Of course, (*b*) represents the explicit starting point of the Kirkwood dipole-induced-dipole formulation of optical properties. Similarly, that approximation (*a*) should be necessary is consistent with the fact that the only way in which many-body effects arise in the effective field theory is through successive

iterations of the *two*-body dipole propagator

$$\mathsf{T}^d(\mathbf{r}) = \frac{3\hat{r}\hat{r} - \mathsf{I}}{r^3} \tag{7.39}$$

the latter (tensor) acting on a dipole moment at an arbitrary point in the fluid to give the electric field (vector) amplitude at a point $\mathbf{r}$ away. More explicitly, solving iteratively the closed integral equation for the effective field in the dielectric, we find

$$\sum_i \alpha \mathbf{E}_{\text{eff}}(\mathbf{R}_i) = \Big\{ N\alpha \mathsf{I} + \alpha^2 \sum_{\substack{\text{all distinct} \\ \text{pairs } ij}} \sum \mathsf{T}^d(\mathbf{R}_i - \mathbf{R}_j)$$

$$+ \alpha^3 \sum_{\substack{\text{all distinct} \\ lmn}} \sum \sum \mathsf{T}^d(\mathbf{R}_l - \mathbf{R}_m) : \mathsf{T}^d(\mathbf{R}_m - \mathbf{R}_n) + \cdots \Big\} \cdot \mathbf{E}_0(R_i) \tag{7.40}$$

Comparing the expansion above with the form of our earlier cluster series

$$\boldsymbol{\alpha}(\mathbf{R}_1, \ldots, \mathbf{R}_N) = \sum_i \boldsymbol{\alpha}^{(1)}(\mathbf{R}_i) + \sum_{i<j} \sum \boldsymbol{\alpha}^{(2)}(\mathbf{R}_i \mathbf{R}_j)$$

$$+ \sum_{l<m<n} \sum \sum {}^{\text{irr}}\boldsymbol{\alpha}(\mathbf{R}_l, \mathbf{R}_m, \mathbf{R}_n) + \cdots$$

the implicit superposition [compare (7.31)] arising in the classical theory is transparent.

Returning now to our "exact" result (7.32) and (7.33), we have

$$\alpha_\perp^{(2)}(r) \equiv {}^{\text{DID}}\alpha_\perp^{(2)}(r) + {}^{\text{CORR}}\alpha_\perp^{(2)}(r)$$

and (7.41)

$$\beta^{(2)}(r) \equiv {}^{\text{DID}}\beta^{(2)}(r) + {}^{\text{CORR}}\beta^{(2)}(r)$$

with

$${}^{\text{DID}}\alpha_\perp^{(2)}(r) = \frac{-2\alpha^2}{r^3} \qquad \text{and} \qquad {}^{\text{DID}}\beta^{(2)}(r) = \frac{+6\alpha^2}{r^3}$$

thereby defining the short-range corrections. In the spirit of our earlier discussion in Section VI, we parameterize the non-DID contributions according to

$${}^{\text{CORR}}\alpha_\perp^{(2)}(r) = \kappa_a e^{\lambda r} \qquad \text{and} \qquad {}^{\text{CORR}}\beta^{(2)}(r) = \kappa_b e^{\lambda r} \tag{7.42}$$

with $\lambda < 0$. Then, substituting (7.41) and (7.42) into (7.33) we have trivially that

$$I|_{\text{"exact"}} = I|_{\text{DID}} + 2I|_{\text{CORR-DID}} + I_{\text{CORR-CORR}} \tag{7.43}$$

Here the first term on the right-hand side is simply the DID limit of (7.33), that is, all $\alpha_\perp$'s and β's in (7.33a) are replaced by their dispersional form [recall the integral appearing in (7.36) and (7.37) where we made

contact with the results of Levine and McQuarrie]; $I|_{\text{CORR-DID}}$ is obtained from (7.33) and (7.33a) by replacing each r-dependent $\alpha_{\perp}^{(2)}$ and $\beta^{(2)}$ in $g(rRs)$ by its DID form and each R-dependent factor by its *short*-range exponential form [the "2" comes from the symmetry of the integrand with respect to interchange of r and R]; finally, the $I_{\text{CORR-CORR}}$ is defined by setting *all* the $\alpha_{\perp}$'s and β's equal to their short-range (*non*-DID) correction form, and can be dropped since it can easily be shown to be negligible compared to the first two terms.

Now, $I|_{\text{DID}}$—the integral defined by (7.33) but with $g(rRs) \to g_{\text{DID}}(rRs)$ [compare the r, R, s integration in (7.36) or (7.37)]—has been evaluated numerically by Levine and McQuarrie[111c] for several different choices of the pair potential energy, and for a wide range of reduced temperatures. Actually, they calculate the quantity

$$2J_5 = \int_0^{\infty} y\,dy \int_0^{\infty} z\,dz \int_{|z-y|}^{z+y} x\,dx e^{-\phi^*(z)/T^*} e^{-\phi^*(y)/T^*} f(x) \frac{3\cos^2\theta - 1}{z^3 y^3} \tag{7.44}$$

where the new, *dimensionless*, variables are defined by: x, y, $z = s/\sigma$, r/σ, R/σ; $\phi^*(y) = u(r)/\varepsilon$; and $T^* = kT/\varepsilon$, where ε is as usual the value of the pair potential $u(r)$ at its minimum and σ is such that $u\,(r = \sigma) = 0$. For hard spheres, the triple integral can be evaluated analytically (see below) and one finds $J_5|_{\text{H.S.}} = -0.104\ (= -5/48)$. In the case of Lennard-Jones $u(r)$, on the other hand, the integrations must be performed numerically and require considerable computation time because of the long-ranged (slowly converging) $1/z^2y^2$ factor: Levine and McQuarrie report[111c] $J_5|_{\text{L.J.}}$ values of -0.105, -0.105, and -0.106 for 6-15, 6-12, and 6-9 potentials, respectively, indicating only about 1% deviation from the hard sphere result. (We have quoted above the Lennard-Jones results for $T^* \approx 30$ since this corresponds to the room temperature helium case in which we are primarily interested.)

The fact that $I|_{\text{DID}}$ is sensibly independent of $u(r)$ suggests that the correction term $I|_{\text{CORR-DID}}$ should also be well-approximated by $u(r) \to u_{\text{H.S.}}(r)$. Returning, then, to the general form of the integral

$$I = (8\pi^2) \int_0^{\infty} r\,dr \int_0^{\infty} R\,dR \int_{|r-R|}^{r+R} s\,ds e^{-\phi(r)} e^{-\phi(R)} f(s) g(rRs) \tag{7.45}$$

and replacing the first three factors in the integrand by their hard sphere limit $H(r - \sigma)H(R - \sigma)[-H(\sigma - s)]$, it is relatively straightforward to show that

$$I \to I|^{\text{H.S.}} = -2(8\pi^2)\left[\int_{\sigma}^{2\sigma} r\,dr \int_{\sigma}^{r} R\,dR \int_{r-R}^{\sigma} s\,ds g(rRs) + \int_{2\sigma}^{\infty} r\,dr \int_{r-\sigma}^{r} R\,dR \int_{r-R}^{\sigma} s\,ds g(rRs)\right] \tag{7.46}$$

Here, of course, σ is just the hard sphere radius. Writing

$$g(rRs) \to g_{\mathrm{DID}}(rRs) = \frac{8\alpha^4}{5r^3R^3}(3\cos^2\theta - 1)$$

and performing explicitly the several successive but trivial integrations involved in (7.46) one finds directly that

$$I\Big|_{\mathrm{DID}}^{\mathrm{H.S.}} = (8\pi^2)\frac{16\alpha^4}{5}J_5^{\mathrm{H.S.}} = -(8\pi^2)\frac{\alpha^4}{3} \tag{7.47}$$

The analytical evaluation of the correction term $I\big|_{\mathrm{CORR\text{-}DID}}$ in the hard sphere limit is just as straightforward, only more tedious. Recall that

$$g_{\mathrm{CORR\ DID}}(rRs) = \Bigg\{ {}^{\mathrm{CORR}}\alpha_{\perp}^{(2)}(r)\ {}^{\mathrm{DID}}\alpha_{\perp}^{(2)}(R) + \frac{{}^{\mathrm{CORR}}\alpha_{\perp}^{(2)}(r)\ {}^{\mathrm{DID}}\beta^{(2)}(R) + {}^{\mathrm{DID}}\alpha_{\perp}^{(2)}(R)\ {}^{\mathrm{CORR}}\beta^{(2)}(r)}{3} + \frac{{}^{\mathrm{CORR}}\beta^{(2)}(r)\ {}^{\mathrm{DID}}\beta^{(2)}(R)}{15}[2\cos^2\theta + 1]\Bigg\} \tag{7.48}$$

with $\cos\theta \equiv (r^2 + R^2 - s^2)/2rR$ as before and the "CORR" and "DID" forms of the polarizability components defined as in (7.41) and (7.42). Now, each term in (7.48) contains *either* ${}^{\mathrm{CORR}}\alpha_{\perp}^{(2)}(r)$ *or* ${}^{\mathrm{CORR}}\beta^{(2)}(r)$, and both of these correction factors—as we discuss below—are essentially zero for $r > 10$ a.u. (i.e., for interparticle distances exceeding about 5 Å the incremental polarizability is well described by the DID theory). Thus, since $2\sigma \approx 10$ a.u. (again, see Section VII.D) we have immediately that [compare (7.46)]

$$I\Big|_{\mathrm{CORR\text{-}DID}}^{\mathrm{H.S.}} \approx -2(8\pi^2)\int_{\sigma}^{2\sigma} r\,dr\int_{\sigma}^{r} R\,dR\int_{r-R}^{\sigma} s\,ds\,g_{\mathrm{CORR\text{-}DID}}(rRs) \tag{7.49}$$

Then, substituting for $g_{\mathrm{CORR\text{-}DID}}$ from (7.48), we find, after considerable manipulation of the elementary integrals involved, that (7.49) can be expressed exactly in the closed form

$$I\Big|_{\mathrm{CORR\text{-}DID}}^{\mathrm{H.S.}} = -\frac{2}{\lambda^3}(8\pi^2)\{Ae^{2\sigma\lambda} - Be^{\sigma\lambda} + C[Ei(2\lambda\sigma) - Ei(\lambda\sigma)]\} \tag{7.50}$$

Here $Ei(x)$ denotes the (tabulated) exponential integral function

$$A = \frac{\alpha^2\kappa_b}{30}\left[\frac{(\sigma\lambda)^5}{36} + \frac{(\sigma\lambda)^4}{72} - \left(2 - \frac{1}{72}\right)(\sigma\lambda)^3 + \frac{16}{3}(\sigma\lambda)^2 + \frac{8}{3}(\sigma\lambda) - \frac{88}{3} + 68\left(\frac{1}{\sigma\lambda}\right) - 80\left(\frac{1}{\sigma\lambda}\right)^2 + 40\left(\frac{1}{\sigma\lambda}\right)^3\right]$$

$$B = \frac{\alpha^2\kappa_b}{30}\left[\frac{(\sigma\lambda)^5}{18} + \frac{(\sigma\lambda)^4}{18} - \frac{35}{9}(\sigma\lambda)^3 + \frac{5}{3}(\sigma\lambda)^2 - \frac{13}{3}(\sigma\lambda) + \frac{16}{3} + 8\left(\frac{1}{\sigma\lambda}\right) - 40\left(\frac{1}{\sigma\lambda}\right)^2 + 40\left(\frac{1}{\sigma\lambda}\right)^3\right] \tag{7.51}$$

and

$$C = \frac{\alpha^2\kappa_b}{30}\left[-\frac{16}{3}(\sigma\lambda)^3 + 4(\sigma\lambda)^4 - \frac{(\sigma\lambda)^6}{18}\right]$$

with κ_b and λ the parameters defining the non-DID incremental polarizability [compare (7.42)].

Thus we have that the "exact" third dielectric virial coefficient can be written as [compare (7.32) and (7.43), etc.]

$$C_d = \frac{8\pi^2}{3}\int_0^\infty r^2\, dr g_1^{\text{H.S.}}(r)\left[\frac{{}^{\text{CORR}}\alpha_{\|}^{(2)}(r) + {}^{\text{CORR}}2\alpha_{\perp}^{(2)}(r)}{3}\right] + \frac{2\pi}{3\alpha}\left[I\big|_{\text{DID}}^{\text{H.S.}} + I\big|_{\text{CORR-DID}}^{\text{H.S.}}\right] + \frac{3}{2\pi\alpha}\left\{\frac{8\pi^2}{3}\int_0^\infty r^2\, dr e^{-\phi_{\text{H.S.}}(r)}\left[\frac{{}^{\text{CORR}}\alpha_{\|}^{(2)}(r) + {}^{\text{CORR}}2\alpha_{\perp}^{(2)}(r)}{3}\right]\right\}^2 \tag{7.52}$$

with $I\big|_{\text{DID}}^{\text{H.S.}}$ and $I\big|_{\text{CORR-DID}}^{\text{H.S.}}$ given by (7.47) and (7.50) and (7.51) and

$${}^{\text{CORR}}\alpha_{\|}^{(2)}(r) \equiv {}^{\text{CORR}}\beta^{(2)}(r) + {}^{\text{CORR}}\alpha_{\perp}^{(2)}(r)$$

the latter two quantities being defined in (7.42). Note that, because of the way in which we have chosen to separate the short- and long-range polarizability effects [compare (7.41) and (7.42)]

$${}^{\text{Exact}}\alpha_{\|}^{(2)} + {}^{\text{Exact}}2\alpha_{\perp}^{(2)} = {}^{\text{CORR}}\alpha_{\|}^{(2)} + {}^{\text{CORR}}2\alpha_{\perp}^{(2)}$$

that is the familiar $\mathcal{O}(\alpha^3/r^6)$ term that appears in the effective field theory [compare (7.35)] has been implicitly lumped into our correction terms. When the ab initio calculations of O'Brien et al. [see Section VI] for the pair incremental polarizability components are fit to [compare (7.41)

and (7.42)]

$$\alpha_{\perp}^{(2)}(r) = \frac{-2a^2}{r^3} + \kappa_a e^{\lambda r} \tag{7.53}$$

and

$$\beta^{(2)}(r) \equiv \alpha_{\parallel}^{(2)}(r) - \alpha_{\perp}^{(2)}(r) = \frac{+6\alpha^2}{r^3} + \kappa_b e^{\lambda r} \tag{7.54}$$

the values $\kappa_a \approx -2.0$ a.u., $\kappa_b \approx -21$ a.u., and $\lambda \equiv -1/r_0$, $r_0 \approx 0.74$ a.u., are obtained. Here we have used the experimental value[106] of 1.39 (a.u.)^3 for the isolated helium atom polarizability α.

Using the results above for $\alpha^{(2)}(r)$ and $\beta^{(2)}(r)$ and taking $\sigma_{\text{H.S.}} = \sigma_{\text{L.J.}} = 4.9$ a.u.[112] gives a calculated value of -7.4 (a.u.)^3 for B_d, agreeing both in sign and in magnitude with the experimental value $-7.4(\pm 5.0)$ a.u. reported by Orcutt and Cole[106]. Here, of course, we have used [compare (7.27*b*)]

$$\begin{aligned} B_d|^{\text{H.S.}} &= \frac{8\pi^2}{3}\int_0^{\infty} r^2\, dr e^{-\phi_{\text{H.S.}}(r)}\left[\alpha_{\perp}^{(2)}(r) + \frac{\beta^{(2)}(r)}{3}\right] \\ &= \frac{8\pi^2}{3}\left(\kappa_a + \frac{\kappa_b}{3}\right)\frac{(2 - 2\sigma\lambda + \sigma\lambda^2)}{(-\lambda^3)}\, e^{\lambda\sigma} \end{aligned} \tag{7.55}$$

Substituting the same values of κ_a, κ_b, λ, and σ into (7.50) and (7.51) we find $I_{\text{CORR-DID}}^{\text{H.S.}} = +2.8$ a.u. Comparing this with the result obtained earlier for $I_{\text{DID}}^{\text{H.S.}}$ [namely, $-8\pi^2\alpha^4/3 \approx -98$ a.u.—compare (7.47)] we see that the short-range polarizability distortion has very little effect on this particular contribution to the third dielectric virial coefficient. Writing (7.52) as

$$C_d = C_d[\text{“}g_1(r)\text{”}] + C_d[\text{“}\alpha^{(3)}\text{”}] + C_d[\text{“}B_d{}^2\text{”}] \tag{7.56}$$

we have, in fact, that

$$C_d[\text{“}\alpha^{(3)}\text{”}] = \frac{2\pi}{3\alpha}\left[I_{\text{DID}}^{\text{H.S.}} + I_{\text{CORR-DID}}^{\text{H.S.}}\right] \approx -143 \text{ a.u.} \tag{7.57}$$

differing only by a few percent from the DID value of -148 a.u.

The *first* term in (7.56),

$$C_d[\text{“}g_1(r)\text{”}] \equiv \frac{8\pi^2}{3}\int_0^{\infty} r^2\, dr g_1^{\text{H.S.}}(r)\left[\alpha_{\perp}^{(2)}(r) + \frac{\beta^{(2)}(r)}{3}\right] \tag{7.58}$$

can be evaluated analytically since, for hard spheres[113]

$$g_1(r) = \begin{cases} 2\pi\left(\frac{3}{2}\sigma^3 - \frac{1}{2}r\sigma^2 + \frac{r^3}{24}\right), & \sigma \leqslant r \leqslant 2\sigma \\ 0, & \text{otherwise} \end{cases} \tag{7.59}$$

TABLE I

Values of the Third Dielectric Virial Coefficient for Helium Calculated with and without Corrections (CORR) Due to Short Range Electronic Distortion, and Broken up into the Various Contributions [Compare (7.56)] Discussed in the Text

	C_d["$g_1(r)$"]	C_d["$\alpha^{(3)}$"]	C_d["$B_d{}^2$"]	C_d*
DID	+61	−148	+0.2	−87
CORR	−949	+5	+19	−925
Total	−888	−143	+19	−1012

* Units of (a.u.)^9/molecule^3 [1 a.u. = 0.529 Å and 1(a.u.)^3/molecule = 0.0891 (cc/mole)].

More explicitly, substituting (7.53) and (7.54), and (7.59) into (7.58), we find

$$C_d[\text{“}g_1(r)\text{”}] = \frac{16\pi^3}{3}\left(\kappa_a + \frac{\kappa_b}{3}\right) f(r_0, \sigma)e^{-\sigma/r_0} + \mathcal{O}(e^{-2\sigma/r_0}) \tag{7.60}$$

where

$$f(r_0, \sigma) = 5r_0{}^6 + 5\sigma r_0{}^5 - \frac{\sigma^2 r_0{}^4}{2} - \frac{5}{6}\sigma^3 r_0{}^2 + \frac{\sigma^4 r_0{}^2}{24} + \frac{5}{24}\sigma^5 r_0 \tag{7.61}$$

Again, using $\kappa_a \approx -2$ a.u., $\kappa_b \approx -21$ a.u., $r_0 \approx 0.74$ a.u., and $\sigma \approx 4.9$ a.u. as discussed above, we have

$$C_d[\text{“}g_1(r)\text{”}] \approx -888 \text{ a.u.} \tag{7.62}$$

as opposed to the value of +61 a.u. obtained in the DID theory [namely, with

$$\alpha^{(2)}(r) + \frac{\beta^{(2)}(r)}{3} \to + \frac{4\alpha^3}{r^6}\Big]$$

Finally, for the last contribution to the third dielectric virial coefficient we have

$$C_d[\text{“}B_d{}^2\text{”}] = \frac{3}{2\pi\alpha}(B_d)^2 \approx 19 \text{ a.u.} \tag{7.63}$$

where we have used the “exact” B_d calculated according to (7.55). Recall that the DID limit of this last term, which is about *a hundred times smaller*, is the one which has been implicitly lumped into the $\mathcal{O}(\alpha^4)$ of (7.37).

The results from directly above are displayed in Table I where we have broken up the three contributions [compare (7.52) and (7.56)] to the third dielectric virial coefficient into their DID and short-range parts. Recalling our earlier results for B_d [compare (7.55)] and A_d [compare

(7.27a)] we have, for gaseous helium at room temperature,

$$\frac{\varepsilon - 1}{\varepsilon + 2}\frac{1}{\rho} = A_d + B_d\rho + C_d\rho^2 + \cdots \tag{7.64}$$

with

$$A_d = 0.519 \text{ cc/mole} \tag{7.64a}$$

$$B_d = -0.060 \text{ (cc/mole)}^2 \tag{7.64b}$$

and

$$C_d = -0.716 \text{ (cc/mole)}^3 \tag{7.64c}$$

[Here we have changed to units used most prevalently in the experimental literature—for example, a.u. $(\text{length})^3/\text{molecule} \rightarrow 0.0891$ cc/mole.] In Fig. 4 the Clausius-Mosotti function given by (7.64) is plotted versus density. We also display there the contrary predictions of the Kirkwood DID theory (see Table I).

The *linear de*crease in $(\varepsilon - 1/\varepsilon + 2)(1/\rho)$ for low densities is precisely what has been observed by Orcutt and Cole.[106] A least-squares fit of the

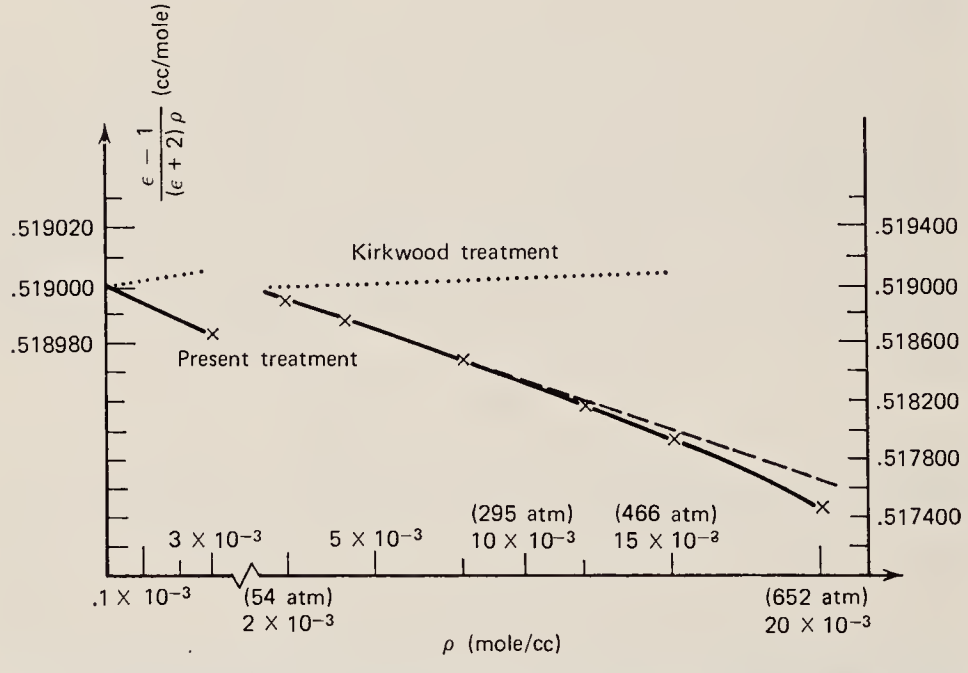


Fig. 4. The Clausius-Mossotti function $(\varepsilon - 1)/(\varepsilon + 2)/\rho$ is plotted here versus density, ρ (the numbers in parentheses indicate the corresponding pressures in atm). The horizontal scale is broken to conveniently display the "low"- and "high"-pressure regimes, the dielectric functions for the latter being indicated by the ordinate scale on the right. The solid line denotes our results including short-range electronic distortion [Compare (7.64)]—the dashed line extrapolation from the linear regime is drawn to point up the quadratic curvature from triplet effects. The dotted line refers to the Kirkwood DID limit.

(less reliable) data up to pressures of 100 atm shows no statistically meaningful trace of a quadratic correction. Our estimate of the third dielectric virial coefficient indicates, in fact, that triplet effects will not be observable until accurate measurements can be performed at gas densities at least three times as high. However, as far as we are aware, no such data are presently available.

Note also the *increasing* (versus ρ) discrepancy (see Fig. 4) between our "exact" Clausius-Mosotti function and the Kirkwood theory. This failure of the DID description has been confirmed and extended into the liquid region by recent machine calculations. More explicitly, Alder and Strauss[114] have computed the thermal average defining the dielectric function in [compare (7.25) and (7.30)]

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi}{3V}\left\langle N\alpha + \frac{N^2}{2!}\alpha^2 T_{ZZ}{}^{d}(\mathbf{R}_1 - \mathbf{R}_2) + \frac{N^3}{3!}T_{ZZ}{}^{d}(\mathbf{R}_1 - \mathbf{R}_2)T_{ZZ}{}^{d}(\mathbf{R}_1 - \mathbf{R}_3)\right\rangle \quad (7.65)$$

The pair and (superposed) triplet quantities are averaged over the *full* (ρ-dependent) two- and three-body equilibrium probabilities (rather than their zero pressure limits as was appropriate for our virial expansion) using the machine method of molecular dynamics. Comparing with the recently compiled data on gas-through-solid *argon*[115], their calculated values for $(\varepsilon - 1/\varepsilon + 3)/(1/\rho)$ are increasingly (with density) too large, this marked deviation arising, of course, from the neglect of short range electronic distortion effects. Recall that we found both B_d and C_d negative for helium, and nearly 10 times larger in magnitude than the corresponding DID values. In argon, where B_d happens to be *positive*, its magnitude is much less than the DID estimate[106]—in both fluids, the effect of the short-range electronic distortion is to *lower* the values of the dielectric function. A similar conclusion has been discussed in previous sections in the case of depolarized light scattering by these same systems.

In summary then, we have presented formally exact expressions, within the Born-Oppenheimer approximation for the separation of electronic and nuclear motions, for the first three coefficients in the density expansion of the dielectric function of a simple fluid. Since these ensemble averages involve trivial integrations over the zero-pressure limits of the 1-, 2-, and 3-particle distribution functions, no assumptions need to be made about the many body structure of the sample. Rather it is necessary to know only the singlet, pair, and triplet irreducible cluster components of the adiabatic ZZ-polarizability $\alpha_{ZZ}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$. This latter quantity assumes the DID form for large interparticle distances, and implicitly includes the

Jansen-Mazur type corrections,[116] as well as all electronic overlap and exchange effects at shorter distances. Trivially, $\alpha_{ZZ}^{(1)}(\mathbf{R}_i)$ is the isolated atom α and is determined directly by low-pressure dielectric measurements and/or electronic structure calculations. (The best values obtained[80] in these ways differ by only a few percent.) The *two*-body incremental polarizability, $\alpha_{ZZ}^{(12)}(\mathbf{R}_{ij})$, in turn has been inferred from ab initio computations[80] and by, say, moment analyses of inelastic depolarized light scattering spectra (see Section VI). We have chosen here to use the O'Brien et al. calculated values for $\alpha^{(12)}(\mathbf{r})$, since they comprise the *only* theoretical result which is qualitatively correct, and therefore invite naturally these self-consistency checks. The only quantity which is *not* "known" in our formulation, then, is the triplet (irreducible cluster component of the) polarizability. Since $\alpha^{(123)}(\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3)_1$ *for large interparticle distances* assumes a simple superposed form involving the (intrinsically two-body) dipole propagators, it has been speculated that a similar form obtains at shorter distances as well but with the $T_{ZZ}{}^{d}(\mathbf{r})$'s replaced by the exact $\alpha_{ZZ}^{(12)}(\mathbf{r})$'s [compare (7.41), (7.42), and (7.31)]. This postulate leads directly to a simple expression for the third dielectric virial coefficient in terms of known (independently determined) quantities.

It is precisely because of the earlier mentioned ambiguities and uncertainties associated with the two-body incremental polarizability components, that we have presented all of the exact results in the form of elementary functions involving the not-well-enough-known parameters κ_a, κ_b, and r_0. [In the case where $\alpha_{\parallel}(r)$ and $\alpha_{\perp}(r)$ cannot be satisfactorily fit to exponentials, one must, of course, use the more general results given by (7.27b), (7.32), (7.33), etc.] In effect we are simply suggesting that the dielectric virial coefficients be used *not* to test intermolecular potentials [assuming

$$\alpha_{ZZ}^{(12)}(\mathbf{r}) \rightarrow {}^{\mathrm{DID}}\alpha_{ZZ}^{(12)}(\mathbf{r}) = 2\alpha^2 T_{ZZ}{}^{d}(\mathbf{r})$$

see, for example, Ref. 111] but rather, more meaningfully, to test ${}^{\mathrm{exact}}\alpha_{ZZ}^{(12)}(\mathbf{r})$'s. This follows from the fact that B_d and C_d—compare (7.27*b*) and (7.32)—are much more sensitive to incremental polarizabilities than to $u(r)$'s.

C. Collision Induced versus Double Scattering in the Gas-Liquid Critical Region

In the final discussion below we return to the classical description of Section V and show that it is possible to separate the contributions to the total depolarization into two distinct kinds: (*1*) single scatterings off anisotropic clusters of interacting particles, and (*2*) *successive* single scatterings off undistorted atoms separated by macroscopic distances.

The theoretical and experimental work referred to up until now has implicitly studied only the first of these mechanisms because essentially, away from the gas-liquid critical point, all of the depolarization arises from *collision-induced* effects characterized by the short-range, many-body structure of the simple fluid. Near enough the critical point, however (as we suggest below), the true *double scattering* events can become dominant and the depolarization assumes a new and dramatic dependence on temperature and sample size (these latter effects arising from the onset of long-range correlations).

An approach similar to that described below has been used[73] to describe the contributions of double scattering to the angle dependence of the integrated *polarized* intensity, $I^{zz}(\phi_s)$. In particular it can be shown that these contributions give rise to convex-downward curvature in what otherwise would be linear, Ornstein-Zernicke-Debye plots ($1/I^{zz}$ versus $\sin^2 \phi_s/2$). These experimentally observable deviations from "classical" critical behavior have been interpreted most commonly in terms of corrections to the long-range pair correlation function, but can be quantitatively reanalyzed instead as small double scattering effects.[73] Consistent with the purposes of this review, however, we discuss below only the integrated (over frequency) *de*polarized intensity, and its angle dependence in the gas-liquid critical region. A calculation of the *in*elastic spectrum (line*shape*) involves essentially the same mathematical and physical approximations, and is treated elsewhere.[73]

Consider again the effective electric field $\mathbf{E}(\mathbf{r}t)$ in our sample composed of N particles in a volume V_s, each characterized by a scalar polarizability α and all in thermal equilibrium at a temperature T. Since we are treating the system in the static approximation, that is, we neglect the time dependence of the density fluctuations (see, however, Ref. 117), a monochromatic incident field

$$\mathbf{E}_0(\mathbf{r}t) = \mathbf{E}_0 e^{i\mathbf{k}_0\cdot\mathbf{r}} e^{-i\omega_0 t}$$

assures that $\mathbf{E}(\mathbf{r}t)$ has the form $\mathbf{E}(\mathbf{r})e^{-i\omega_0 t}$. Writing as before [see Section V] the effective electric field at $\mathbf{r}t$ as the sum of the incident field $\mathbf{E}_0(\mathbf{r}t)$ and the dipolar fields arising from the moments $\alpha\rho(\mathbf{r}')\mathbf{E}(\mathbf{r}'t)$ induced at all neighboring points, we have directly that

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0 e^{i\mathbf{k}_0\cdot\mathbf{r}} + \alpha\int_{V_I} d\mathbf{r}_2 \mathsf{T}_{k_0}(\mathbf{r}-\mathbf{r}_2)\cdot\rho(\mathbf{r}_2)\mathbf{E}_0 e^{i\mathbf{k}_0\cdot\mathbf{r}_2}$$

$$+ \alpha^2\int_{V_s-\sigma_{\mathbf{r}_2}} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \mathsf{T}_{k_0}(\mathbf{r}-\mathbf{r}_1)\cdot\rho(\mathbf{r}_1)\mathsf{T}_{k_0}(\mathbf{r}_1-\mathbf{r}_2)$$

$$\times \rho(\mathbf{r}_2)\mathbf{E}_0 e^{i\mathbf{k}_0\cdot\mathbf{r}_2} + \mathcal{O}(\alpha^3) \quad (7.66)$$

Here, as explained in Section V, $\rho(\mathbf{r}') = \sum_{i=1}^{N}\delta(\mathbf{r}' - \mathbf{R}_i)$ is the instantaneous number density at $\mathbf{r}'$ corresponding to the fluid configuration $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ and

$$\mathsf{T}_{k_0}(\mathbf{r}) \equiv e^{ik_0 r}\left(\frac{3}{r^3} - \frac{3ik_0}{r^2} - \frac{{k_0}^2}{r}\right)\hat{r}\hat{r} + e^{ik_0 r}\left(\frac{k_0}{r} + \frac{ik_0}{r^2} - \frac{1}{r^3}\right)\mathsf{I} \quad (7.67)$$

acts on an oscillating ($\omega_0 = ck_0$) dipole moment to give the resulting electric field at a vector distance $\mathbf{r}$ away. The subscript I on the integration volume in the α-term above refers to the fact that the only contributions to the effective field in this order arise from the dipole moments induced in the illuminated volume. In the α^2 terms, on the other hand, the outer integration extends over the entire sample (excluding only an atomic volume about $\mathbf{r}_2$) and includes re-radiation effects from particles polarized in the "dark" volume where the incident field vanishes, that is, this second-order term contains double scatterings (compare the "1/r" terms in the dipole propagators) as well as collision-induced effects in the "near (static) zone"[118], involving $|\mathbf{r}_1 - \mathbf{r}_2| \ll 1/k_0$. More explicitly, since we can exactly rewrite (7.67) as

$$\mathbf{E}(\text{at } \mathbf{r}, \text{ due to dipole } \boldsymbol{\rho} \text{ at the origin}) = e^{ik_0 r}\left(\frac{{k_0}^2}{r}\right)[(\hat{r} \times \boldsymbol{\rho}) \times \hat{r}]$$

$$+ e^{ik_0 r}\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)[3\hat{r}(\hat{r} \cdot \boldsymbol{\rho}) - \boldsymbol{\rho}] \quad (7.68)$$

the transverse nature of the electric field in the "far ($k_0 r \gg 1$) zone" becomes transparent. Thus the $({\omega_0}^2/c^2)(e^{ik_0 r}/r)\hat{r}\hat{r}$ term in (7.67) (the I-component can be dropped since it survives only in the case of the polarized spectrum) arises from a true *radiation* term, allowing the new long-range correlation effects which we discuss below to be interpreted as real double scattering contributions, as opposed to the familiar collision-induced processes involving the shorter-ranged "$1/r^3$, $1/r^2$" that are dominant *away* from the critical region. Indeed, in all the actual calculations of the depolarized intensity treated up until now, *only* the ($k_0 r \ll 1$) static limit of the dipole propagator was considered, consistent with our explicit demonstration below that for $|T - T_c|$ greater than a degree or so, only the shortest-range correlation effects need be included. This distinction will be pursued in detail throughout the rest of our discussion.

We refer again to the experimental scattering geometry shown in Fig. 1. Incident light of frequency ω_0 propagates along the negative x-axis ($\mathbf{k}_0 = -k_0\hat{x}$) and is linearly polarized in the z-direction ($\mathbf{E}_0 = E_0\hat{z}$). A detector on the y-axis at some point $\mathbf{R} = R\hat{y}$ far removed from the

sample measures the integrated intensity $I_{\mathbf{R}}{}^{ZX}$ of the x-component of the scattered light. Substituting for $\mathbf{E}(\mathbf{R})$ from (7.66) and using [for $|\mathbf{R}| \gg |\mathbf{r}_1|$]

$$\mathsf{T}_{k_0}(\mathbf{R} - \mathbf{r}_1) \approx \frac{k_0{}^2 e^{ik_0(|\mathbf{R}| - \mathbf{r}_1 \cdot \hat{R})}}{|\mathbf{R}|} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.67}$$

we have directly that (here $\mathbf{k}' \equiv k_0 \hat{R}$, the propagation vector of the scattered light)

$$\begin{aligned} I_{\mathbf{R}}{}^{ZX} = \frac{\alpha^4}{R^2} k_0{}^4 |E_0|^2 \int_{V_s - \sigma_{\mathbf{r}_2}} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \int_{V_s - \sigma_{\mathbf{r}_4}} d\mathbf{r}_3 \int_{V_I} d\mathbf{r}_4 \\ \times \exp\left[i\mathbf{k}_0 \cdot (\mathbf{r}_2 - \mathbf{r}_4) - i\mathbf{k}' \cdot (\mathbf{r}_1 - \mathbf{r}_3)\right] [\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ} \\ \times [\mathsf{T}_{k_0}(\mathbf{r}_3 - \mathbf{r}_4)]_{XZ}{}^* \langle \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rho(\mathbf{r}_3)\rho(\mathbf{r}_4)\rangle \end{aligned} \tag{7.68}$$

Note that this result is identical to (5.16) except that now we are careful to label explicitly the V_I and V_s integration volumes. The ensemble average $n^{(4)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)$ appearing in (7.68) can be written in terms of the usual n-particle conditional distribution functions according to

$$\begin{aligned} &\langle \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rho(\mathbf{r}_3)\rho(\mathbf{r}_4)\rangle \\ &= \delta_{13}\delta_{24}\rho^2 g^{(2)}(12) + \delta_{14}\delta_{23}\rho^2 g^{(2)}(12) + \delta_{24}\rho^3 g^{(3)}(123) + \delta_{14}\rho^3 g^{(3)}(123) \\ &\quad + \delta_{23}\rho^3 g^{(3)}(124) + \delta_{13}\rho^3 g^{(3)}(124) + \rho^4 g^{(4)}(1234) \end{aligned} \tag{7.69}$$

Note that terms containing $\delta_{12} \equiv \delta(\mathbf{r}_1 - \mathbf{r}_2)$ or $\delta_{34} \equiv \delta(\mathbf{r}_3 - \mathbf{r}_4)$ have been dropped because of the integration restrictions $\mathbf{r}_1 \neq \mathbf{r}_2$, $\mathbf{r}_3 \neq \mathbf{r}_4$. Our basic result for the depolarized intensity then becomes

$$\begin{aligned} I_{\mathbf{R}}{}^{ZX} = \frac{\alpha^4 k_0{}^4 |E_0|^2}{R^2} \Bigg\{ &\rho^2 \int_{V_s} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \, |\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)|_{XZ}{}^2 \, g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\ &+ \rho^2 \int_{V_I} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \exp\left[i(\mathbf{k}_0 + \mathbf{k}') \cdot (\mathbf{r}_2 - \mathbf{r}_1)\right] \\ &\times |\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)|_{XZ}{}^2 \, g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\ &+ \rho^3 \int_{V_s} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \int_{V_s} d\mathbf{r}_3 \exp\left[-i\mathbf{k}' \cdot (\mathbf{r}_1 - \mathbf{r}_3)\right] [\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ} \\ &\times [\mathsf{T}_{k_0}(\mathbf{r}_2 - \mathbf{r}_3)]_{XZ}{}^* g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \\ &+ \rho^3 \int_{V_I} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \int_{V_s} d\mathbf{r}_3 \exp\left[i\mathbf{k}_0 \cdot (\mathbf{r}_1 - \mathbf{r}_2) + i\mathbf{k}' \cdot (\mathbf{r}_3 - \mathbf{r}_2)\right] \end{aligned}$$

$$\times [\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ}[\mathsf{T}_{k_0}(\mathbf{r}_2 - \mathbf{r}_3)]_{XZ}{}^* g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$$

$$+ \rho^3 \int_{V_S} d\mathbf{r}_1 \int_{V_I} dr_2 \int_{V_I} d\mathbf{r}_3 \exp [i\mathbf{k}_0 \cdot (\mathbf{r}_2 - \mathbf{r}_3) + i\mathbf{k}' \cdot (\mathbf{r}_2 - \mathbf{r}_1)]$$

$$\times [\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ}[\mathsf{T}_{k_0}(\mathbf{r}_2 - \mathbf{r}_3)]_{XZ}{}^* g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$$

$$+ \rho^3 \int_{V_I} d\mathbf{r}_1 \int_{V_S} d\mathbf{r}_2 \int_{V_I} d\mathbf{r}_3 \exp [i\mathbf{k}_0 \cdot (\mathbf{r}_1 - \mathbf{r}_3)][\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ}$$

$$\times [\mathsf{T}_{k_0}(\mathbf{r}_2 - \mathbf{r}_3)]_{XZ}{}^* g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$$

$$+ \rho^4 \int_{V_S} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \int_{V_S} d\mathbf{r}_3 \int_{V_I} d\mathbf{r}_4$$

$$\times \exp [i\mathbf{k}_0 \cdot (\mathbf{r}_2 - \mathbf{r}_4) - i\mathbf{k}' \cdot (\mathbf{r}_1 - \mathbf{r}_3)]$$

$$\times [\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ}[\mathsf{T}_{k_0}(\mathbf{r}_3 - \mathbf{r}_4)]_{XZ}{}^* g^{(4)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) \Big\} \qquad (7.70)$$

Equation (7.70), *away from the critical point*, has been extensively studied by many workers, as discussed in the preceding sections of this review. In this case $g^{(2)}$, $g^{(3)}$, and $g^{(4)}$ fall off rapidly (to unity) with interatomic separations, and the exponentials can be replaced by 1 for the smaller distances of interest. The short-range form of the dipole propagator can also be used:

$$\mathsf{T}_{k_0}(\mathbf{r}) \approx \frac{3}{r^3}\hat{r}\hat{r} - \frac{1}{r^3}\mathsf{I} \qquad \text{for} \qquad r \ll \frac{1}{k_0} \qquad (7.71)$$

Under these conditions (7.70) reduces precisely to (6.11) which has been studied: for moderate pressure gases through the ρ^3-terms, using a superposition approximation for $g^{(3)}$, and for denser fluids and liquids using the method of molecular dynamics. As explained in Section VI these investigations suggest that the $g^{(3)}$ and $g^{(4)}$ terms provide significant negative contributions to the total intensity for sufficiently dense fluids; they reduce the predicted intensity by an order of magnitude from what would be obtained by simple extrapolation of the dilute gas (binary collision) scattering intensities.

As also discussed above, the breakdown of the dipole-induced-dipole (DID) description of collision-induced depolarization due to its implicit neglect of short-range polarizability distortion (electronic overlap, exchange, etc.) can in itself be of considerable significance. These latter effects further decrease the integrated scattering intensities and are quantitatively as important as the many-body structural correlations. We have also shown (Section VII.B) how non-DID polarization processes can account for the observed density dependence of dielectric functions and

Kerr (electric-field-induced birefringence) coefficients in these same simple fluids. Since we concentrate below, however, on the *long*-range correlation contributions to the integrated scattering intensity, it is appropriate to neglect electronic overlap and exchange effects, and so on, picking them up only at the end where we normalize everything with respect to the collision-induced "background."

Suppose we now evaluate the integrals in (7.70) near the critical point. The $g^{(2)}$ terms are straightforward. It is convenient to rewrite these integrals in terms of the Ursell functions,

$$u(r) \equiv u^{(2)}(r) = g^{(2)}(r) - 1 \tag{7.72}$$

and to break up $u(r)$ into two parts:

$$u(r) = \begin{cases} u_{\mathrm{SR}}(r) & r < \bar{r} \\ u_{\mathrm{OZ}}(r) & r \geqslant \bar{r} \end{cases} \tag{7.73}$$

Here u_{OZ} is the Ornstein-Zernike form for $u(r)$,

$$u_{\mathrm{OZ}}(r) = \frac{\xi_0}{r} e^{-r/\xi} \tag{7.74}$$

ξ is the correlation length, and $\bar{r}$ is the distance at which the Ornstein-Zernike form for $u(r)$ becomes valid—it is somewhat arbitrary, but will not turn out to be very important in the results. (We have in mind an $\bar{r}$ of 5 to 20 Å.) Here $u_{\mathrm{SR}}(r)$ is the short-range part of $u(r)$, which is dominated by the nearest-neighbor peak, and which is not expected to change very much with temperature in the critical region.

With these definitions the $g^{(2)}$ terms from (7.70) become (aside from the constant $\alpha^4 k_0^4 |E_0|^2/R^2$ in front)

$$\begin{aligned} &\rho^2 V_I \int_{V_s - \sigma_0} d\mathbf{r}\, |\mathsf{T}_{k_0}(\mathbf{r})|_{XZ}{}^2 + \rho^2 V_I \int_{V_I - \sigma_0} d\mathbf{r} \exp\left(i(\mathbf{k}_0 + \mathbf{k}') \cdot \mathbf{r}\right) |\mathsf{T}_{k_0}(\mathbf{r})|_{XZ}{}^2 \\ &\quad + \rho^2 V_I \int_{V_s - \bar{r}_0} d\mathbf{r}(1 + \exp\left[i(\mathbf{k}_0 + \mathbf{k}') \cdot \mathbf{r}\right]) |\mathsf{T}_{k_0}(\mathbf{r})|_{XZ}{}^2 u_{\mathrm{OZ}}(r) \\ &\quad + \rho_2 V_I \int_{V_s - \sigma_0} d\mathbf{r}(2)\left(\frac{9}{r^6}\right)(\hat{r} \cdot \hat{x})^2(\hat{r} \cdot \hat{z})^2 u_{\mathrm{SR}}(r) \end{aligned} \tag{7.75}$$

Here we have changed the variable of integration to $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ (implicitly assuming that $V_s \gg V_I$), and have used the short-range nature of U_{SR} to replace the dipole propagator by its small distance form [see (7.71)]. The notation $\bar{r}_0$ refers to a spherical volume of radius $\bar{r}$ (centered at the origin **0**) which is excluded from the **r**-integration. The integrals in

(7.75) can be calculated exactly, the dominant terms being given by

$$g^{(2)}\text{ term} = \rho^2 V_I\left[\frac{8\pi}{5\sigma^3} + \frac{24\pi}{5}\int_\sigma^{\bar{r}} dr\,\frac{u_{\mathrm{SR}}(r)}{r^4} + \frac{6\pi\xi_0}{5\bar{r}^4} + \frac{4\pi k_0^4 R_s}{15}\right] \tag{7.76}$$

where σ is an atomic radius, and R_s is a linear dimension characterizing the sample size. It is easy to show (see below) that the first two terms in (7.76) are orders of magnitude larger than the last two, so that the $g^{(2)}$ contributions to ${I_{\mathrm{R}}}^{ZX}$ are determined by short-range, collision-induced effects.

The Ursell expansion for $g^{(3)}$ is

$$g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = 1 + u(r_{12}) + u(r_{13}) + u(r_{23}) + u^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \tag{7.77}$$

but, for reasons of symmetry, the only nonzero contributions to (7.70) from $g^{(3)}$ come from $u(r_{13}) + u^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$. First, we restrict our attention to $u(r_{13})$, and discuss $u^{(3)}$ in Section V. We start by examining the $u(r_{13})$ term arising from the first of the four $g^{(3)}$ terms in ${I_{\mathrm{R}}}^{ZX}$:

$$\begin{aligned}\rho^3\int_{V_f} d\mathbf{r}_1\int_{V_I} d\mathbf{r}_2\int_{V_s} d\mathbf{r}_3 \exp\,[i\mathbf{k}'\cdot(\mathbf{r}_3 - \mathbf{r}_1)][\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)]_{XZ} \\ \times [\mathsf{T}_{k_0}(\mathbf{r}_2 - \mathbf{r}_3)]_{XZ}{}^* u(\mathbf{r} - \mathbf{r}_1) \\ = \rho^3 V_I\int_{V_s} d\mathbf{r}_{12}\int_{V_s} d\mathbf{r}_{23}\exp\,[i\mathbf{k}'\cdot(\mathbf{r}_{12} + \mathbf{r}_{23})] \\ \times [\mathsf{T}_{k_0}(\mathbf{r}_{12})]_{XZ}[\mathsf{T}_{k_0}(\mathbf{r}_{23})]_{XZ}{}^* u(\mathbf{r}_{12} + \mathbf{r}_{23})\end{aligned} \tag{7.78}$$

This is the only $g^{(3)}$ term considered by Mountain [64] in his earlier discussion of depolarized light scattering near the gas-liquid critical point. As we shall show it is much larger than the other three in (7.70).

In order to "decouple" the six-fold integration appearing in (7.78) it is convenient to express $u(r)$ as the Fourier transform of the static structure factor:

$$\begin{aligned}u(r) &\equiv \frac{1}{(2\pi)^3\rho}\int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}}[S(k) - 1] \\ &= \frac{1}{(2\pi)^3}\int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}}\left[\frac{4\pi\xi_0}{k^2 + (1/\xi)^2} + \frac{S_{\mathrm{SR}}(k)}{\rho}\right]\end{aligned} \tag{7.79}$$

where $S_{\mathrm{SR}}(k)$ is the contribution to $S(k)$ from the short-range part of $u(r)$. Putting (7.79) into (7.78) and reversing the order of integration,

(7.78) assumes the convolution form

$$\frac{\rho^3 V_I}{(2\pi)^3}\int d\mathbf{k}\,|g(\mathbf{k}'-\mathbf{k})|^2\left[\frac{4\pi\xi_0}{k^2+(1/\xi)^2}+\frac{S_{\mathrm{SR}}(\mathbf{k})}{\rho}\right] \tag{7.80}$$

where

$$g(\mathbf{k}_1)=\int_{V_s} d\mathbf{r}e^{i\mathbf{k}_1\cdot\mathbf{r}}[\mathsf{T}_{k_0}(\mathbf{r})]_{XZ} \tag{7.81}$$

$|g(\mathbf{k}_1)|^2$ is calculated in Appendix D, where we show that

$$|g(\mathbf{k}_1)|^2=(\hat{k}_1\cdot\hat{x})^2(\hat{k}_1\cdot\hat{z})^2[h(k_1)+8\pi^3k_0{}^2R_s\delta(k_1-k_0)] \tag{7.82}$$

with R_s again the sample "radius" and $h(k_1)$ given by (A.27) and (A.28). Substituting (7.82) into (7.80) gives

$$\frac{\rho^3V_I}{(2\pi)^3}\int dk_1(\hat{k}_1\cdot\hat{x})^2(\hat{k}_1\cdot\hat{z})^2\left\{\frac{32\pi^4\xi_0k_0{}^2R_s\delta(k_1-k_0)}{|\mathbf{k}'-\mathbf{k}_1|^2+(1/\xi)^2}+\frac{4\pi\xi_0h(k_1)}{|\mathbf{k}'-\mathbf{k}_1|^2+(1/\xi)^2}\right.$$
$$\left.+\frac{8\pi^3k_0{}^2R_s}{\rho}\delta(k_1-k_0)S_{\mathrm{SR}}(|\mathbf{k}'-\mathbf{k}_1|)+\frac{h(k_1)}{\rho}S_{\mathrm{SR}}(|\mathbf{k}'-\mathbf{k}_1|)\right\} \tag{7.83}$$

for the particular $g^{(3)}$ contribution to $I_{\mathrm{R}}{}^{ZX}$ which we are considering here.

The integration over the first of the four terms in (7.83) can be evaluated trivially. The second has been performed[117] numerically and is found to be smaller than the first term by a factor of 10^5 even for a sample radius as small as 0.1 mm. The third term is also negligible since $S_{\mathrm{SR}}(k)$ is much smaller than $S_{\mathrm{OZ}}(k)$ for k in the optical wavenumber range. The final integration contains the short-range contribution to $g^{(3)}$: it is expected to be insensitive to the temperature near T_c and will be somewhat smaller than the short range $g^{(2)}$ term and opposite in sign. The effect of this last term can therefore be grouped with the short-range $g^{(2)}$ term as an overall collision-induced contribution.

Equation (7.83) then becomes, after integrating over $\mathbf{k}_1$

$$g^{(3)}\text{ term}=4\pi^2\rho^3V_I\xi_0R_sk_0{}^2\left[\frac{2}{3a}-\frac{a^2+1}{4a^3}+\frac{(1-a^2)^2}{8a^4}\ln\left(\frac{1+a}{1-a}\right)\right] \tag{7.84}$$

where $a=2k_0{}^2/[2k_0{}^2+(1/\xi)^2]$. For $k_0\xi\lesssim 0.3$ this reduces to

$$g^{(3)}\text{ term}\approx\frac{16\pi^2}{15}\rho^3V_I\xi_0R_sk_0{}^2(k_0\xi)^2 \tag{7.85}$$

while for $k_0\xi\gg 1$, it approaches asymptotically

$$\tfrac{2}{3}\pi^2\rho^3V_I\xi_0R_sk_0{}^2 \tag{7.86}$$

As shown in Appendix D, the other three $g^{(3)}$ terms in (7.70) do not have this R_s volume dependence, and so are negligible.

The $g^{(4)}$ term can also be written in an Ursell expansion (as the sum of 1, six $u^{(2)}$'s, three products of pairs of $u^{(2)}$'s, four $u^{(3)}$'s, and $u^{(4)}$). For symmetry reasons the only nonzero contributions to (7.70) come from

$$u(\mathbf{r}_{13})u(\mathbf{r}_{24}) + u(\mathbf{r}_{14})u(\mathbf{r}_{23}) + u^{(4)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) \tag{7.87}$$

The contribution from the first term is

$$\rho^4 \int_{V_s} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \int_{V_s} d\mathbf{r}_3 \int_{V_I} d\mathbf{r}_4 \exp\left[i\mathbf{k}_0 \cdot (\mathbf{r}_2 - \mathbf{r}_4) + i\mathbf{k}' \cdot (\mathbf{r}_3 - \mathbf{r}_1)\right]$$
$$\times \left[\mathsf{T}_{k_0}(\mathbf{r}_1 - \mathbf{r}_2)\right]_{XZ} \left[\mathsf{T}_{k_0}(\mathbf{r}_3 - \mathbf{r}_4)\right]_{XZ}{}^* u(\mathbf{r}_3 - \mathbf{r}_1) u(\mathbf{r}_4 - \mathbf{r}_2) \tag{7.88}$$

Changing variables to $\mathbf{r}_{12}$, $\mathbf{r}_{34}$, $\mathbf{r}_2$, and $\mathbf{r}_4$, we write both $u(r)$'s as Fourier transforms of the static structure factor [compare (7.79)] and perform the trivial integrations over $\mathbf{r}_2$ and $\mathbf{r}_4$. The result is

$$\frac{\rho^4 V_I}{(2\pi)^3} \int d\mathbf{k}_1 \, |g(\mathbf{k}_1)|^2 \left[\frac{4\pi\xi_0}{|\mathbf{k}_1 + \mathbf{k}'|^2 + (1/\xi)^2} + \frac{1}{\rho} S_{\mathrm{SR}}(|\mathbf{k}_1 + \mathbf{k}'|)\right]$$
$$\times \left[\frac{4\pi\xi_0}{|\mathbf{k}_1 + \mathbf{k}_0|^2 + (1/\xi)^2} + \frac{1}{\rho} S_{\mathrm{SR}}(|\mathbf{k}_1 + \mathbf{k}_0|)\right] \tag{7.89}$$

with $g(k_1)$ defined by (7.81) and (7.82) and (A.27) and (A.28).

Eight terms result when (7.89) is multiplied out. The product $h(k_1) \times (S_{\mathrm{SR}}(\mathbf{k}_1 + \mathbf{k}')S_{\mathrm{SR}}(\mathbf{k}_1 + \mathbf{k}_0)$ corresponds to the short-range correlation contribution and can be grouped with the $g^{(2)}$ and $g^{(3)}$ collision-induced terms. Of the other seven terms, six are negligible for the same reasons as discussed for the R_s-independent $g^{(3)}$ terms. After dropping these terms, then, and integrating over k_1, (7.89) becomes

$$4\pi^2 \xi_0{}^2 R_s V_I \rho^4 a^2 \int_0^{\pi} d\theta_1 \int_0^{2\pi} d\phi_1 \frac{\sin^3\theta_1 \cos^2\theta_1 \sin^2\phi_1}{[1 + a\sin\theta_1\cos\phi_1][1 - a\sin\theta_1\sin\phi_1]} \tag{7.90}$$

where $a = 2k_0{}^2/[2k_0{}^2 + (1/\xi)^2]$ as before. This result can be evaluated numerically for arbitrary values of $k_0\xi$, and is graphed in Fig. 5. For $k_0\xi \lesssim 0.3$ it assumes the simple form

$$g^{(4)} \text{ term} \approx \frac{64\pi^3}{15} \rho^4 V_I \xi_0{}^2 R_s (k_0\xi)^4 \tag{7.91}$$

while for $k_0\xi \gg 1$, it approaches

$$16\pi^3(0.193)\rho^4 V_I \xi_0{}^2 R_s \tag{7.92}$$

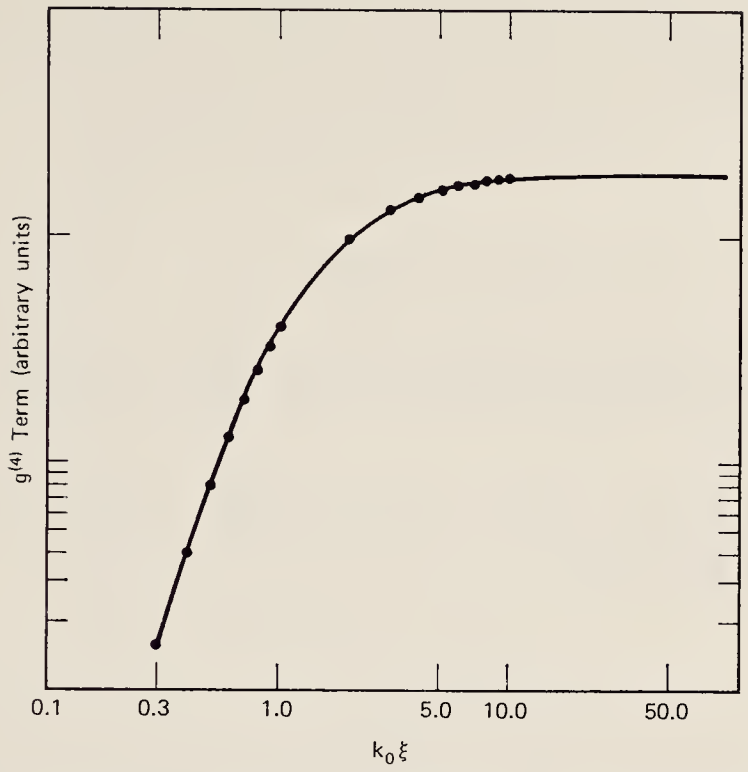


Fig. 5. Dependence on $k_0\xi$ of the "$g^{(4)}$" contribution to the integrated depolarized scattering intensity I^{ZX}. Here ξ is the long-range (Ornstein-Zernike) correlation length and $k_0 = 2\pi/\lambda$ with λ the optical wavelength of the incident light.

The $g^{(4)}$ term from $u(r_{14})u(r_{23})$—compare (7.87)—does not have this volume dependence and is negligible in comparison; the contribution from the $u^{(4)}$-component is discussed below.

It is also possible to perform calculations for scattering angles other than 90°. If we let $\mathbf{k}'$ lie in the xy-plane at an angle ϕ_s with respect to the negative x-axis (for forward and right angle scattering, $\phi_s = 0$ and 90°, respectively) we can calculate the angle-dependence of the scattered light polarized perpendicular to both $\hat{z}$ and $\mathbf{k}'$. We discuss here only the $g^{(4)}$ term, since we show below that it dominates the $g^{(3)}$ term as the critical point is approached. [The short-range correlation ("collision-induced") contributions have negligible angle dependence.] When ϕ_s differs from 90°, (7.90) must be replaced by

$$4\pi^2\xi_0{}^2R_sV_I\rho^4a^2\int_0^{\pi}d\theta_1\int_0^{2\pi}d\phi_1 \times \frac{\sin^2\theta_1\cos^2\theta_1\sin^2\phi_1}{[1 + a\sin\theta_1\cos\phi_1][1 + a\sin\theta_1\cos(\phi_1 + \phi_s)]} \tag{7.93}$$

This integral can be converted to a convergent infinite series and summed numerically for arbitrary $k_0\xi$ and ϕ_s. For $k_0\xi \lesssim 0.3$, we retain terms up to $\mathcal{O}(a^4)$ and find

$$\frac{16\pi^3}{15}\rho^4 V_I \xi_0^2 R_s\left[a^2 + \frac{a^4}{7}(4 + \cos\phi_s - 2\cos^2\phi_s) + \mathcal{O}(a^6)\right]$$

$$\approx \frac{64\pi^3}{15}\rho^4 V_I \xi_0^2 R_s (k_0\xi)^4\left[1 + \frac{4}{7}(k_0\xi)^4(4 + \cos\phi_s - 2\cos^2\phi_s)\right] \quad (7.94)$$

Thus we see that angle dependence should be observable for small enough wavelengths and temperatures close enough to the critical point [so that $(k_0\xi)^4 > 0.001$], although it will not be as dramatic as in the polarized case.

In order to present some numerical estimates of the relative magnitudes of the terms calculated above, we apply the theory to xenon at its critical point, for which we use

$$\rho_c = 5.1 \times 10^{21} \text{ molecules/cc (Ref. 120)}$$
$$\sigma = 3.7 \text{ Å (Ref. 121)}$$
$$\xi_0 = 0.65 \text{ Å (Ref. 122)}$$

We also take $k_0 = 10^5 \text{ cm}^{-1}$. Recall the results given in (7.76) for the $g^{(2)}$ contributions to I_R^{ZX}: the fourth term is negligible even for a huge sample volume. Also, even if we take $\bar{r}$ as small as 2σ, the third term is smaller than the first by a factor of 10. Because of the r^{-4}-dependence in the second term, it is determined almost entirely by the shape and height of the nearest neighbor peak in $g(r)$, which we do not expect to be rapidly changing along the critical isochore. Using the X-ray scattering data for *argon* of Mikolaj and Pings[123] at ρ_c and at $T = -120°\text{C}$ ($T_c = -122.29°\text{C}$), we have calculated this term to be about five times smaller than the first term. Comparable data for xenon have not been published, but similar relative magnitudes should obtain.

The $g^{(2)}$ term discussed above is representative, at least in order of magnitude, of the total collision-induced contribution to the depolarized scattering intensity: at critical densities, the short-range correlation contributions from $g^{(3)}$ and $g^{(4)}$ will be somewhat smaller. As mentioned many times earlier, an accurate calculation would of course have to include the effects of exchange and electronic overlap between atoms, which are neglected in the dipole-induced-dipole approximation; these latter contributions have been shown to lower further the integrated scattering intensities. Alternatively, this short-range contribution could be observed experimentally by studying the depolarized intensity at the critical density but at temperatures away from T_c.

TABLE II

Relative Contributions of the (Short-Range) $g^{(2)}$ Term and the (Long-Range) $g^{(3)}$ and $g^{(4)}$ Terms to the Integrated Depolarized Intensity, for Two Different Sample Cell Radii

$R_s = 1$ cm $\frac{T_c - T^*}{T_c}$	ξ	$g^{(2)}$	$g^{(3)}$	$g^{(4)}$
1.9×10^{-2}	30 Å	1	0.03	1.2
2.4×10^{-3}	100 Å	1	0.3	120
3.6×10^{-4}	300 Å	1	3	12,000
$R_s = 0.1$ mm				
1.9×10^{-2}	30 Å	1	3×10^{-4}	0.012
2.4×10^{-3}	100 Å	1	3×10^{-3}	1.2
3.6×10^{-4}	300 Å	1	0.03	120

* Calculated from the data of Ref. 120a along the critical isochore.

The *long*-range correlation contributions from $g^{(3)}$ and $g^{(4)}$ are given by (7.84) and (7.90) and are easily calculated.

In Table II the relative magnitudes of these different terms are compared for several values of the sample radius, R_s, and the temperature. The $g^{(2)}$ short-range correlation (collision-induced) contributions, which are independent of R_s and ξ, have been lumped under "$g^{(2)}$" and have been normalized to 1; as mentioned earlier the corrections due to $g^{(3)}$ and $g^{(4)}$ short-range effects are not significant for our present purposes.

As Table II indicates, for large enough sample volumes, short enough wavelengths, and temperatures close enough to T_c, long-range effects dominate collision-induced effects. The depolarized intensity, I_R^{XZ}, should then vary as $1/\lambda^8$ rather than as $1/\lambda^4$ (where λ is the wavelength of the incident light) and should be proportional to R_s (the linear dimension characterizing the effective cell size) and ξ^4. Recalling the well-known behavior $I^{ZZ} \sim \xi^2$, the depolarization *ratio* is predicted in this case to increase as ξ^2, although it will still be small. These results are only valid for $k_0\xi \lesssim 0.3$; for larger values of ξ, the $g^{(3)}$ and $g^{(4)}$ terms approach constants, and the total intensity should behave as in Fig. 5. Our prediction[117] that I^{ZX}/I^{ZZ} should *increase* as ξ^2 for $R_s \approx 1$ mm and 80 Å $\lesssim \xi \lesssim$ 300 Å (see Table II) is in good agreement with preliminary xenon data obtained by Swinney (private communication). In particular, his experiments show that the depolarization ratio increases by a factor of ≈ 6 as $\Delta T \equiv T - T_c$ is reduced from ≈ 0.3°K to 0.05°K along the critical isochore (in a sample cell characterized by dimensions on the order of a mm).

In our discussion of sample-cell geometry, we have assumed (for reasons of mathematical tractability) a rather idealized experimental setup, in which a small illuminated volume is enclosed within a spherical sample volume. A more realistic geometry should still agree with our results at least in order of magnitude, and the ξ^4 temperature dependence should still be present. R_s would no longer be the radius of a sphere, but rather a characteristic linear dimension of the sample cell. Also, in most experiments on (polarized) light scattering near the critical point, the collection optics are arranged in such a way that the detector receives light from only a small portion of the sample. In this case the "sample volume" in our theory simply corresponds to the volume actually observed by the detector. By changing the collection optics, one would change R_s. Of course, if "volume dependence" is eliminated, the interesting critical point effects in the depolarized intensity would similarly disappear. Again, our theoretical results for the relationship between the depolarized scattering intensity and the sample size are in qualitative agreement with unpublished data of Swinney (private communication). More explicitly, for each value of the temperature, the experimental values for I^{ZX}/I^{ZZ} show a roughly linear dependence on the (effective) cell dimension which is varied by changing the aperture diameter defining the collection optics.

Throughout our discussion we have implicitly assumed that the correlation length ξ is constant throughout the sample cell. Near the critical point, it is well-known that density gradients become pronounced, with the result that the correlation length varies with height in the cell, reaching its maximum in the middle, at the meniscus. This effect is equivalent to a decrease in the effective sample volume V_s, a correction that should be calculable.

A rather different treatment (from the above) of the depolarized light scattering near the critical point has been suggested by Mountain.[64] He looked at only the integral given in (7.78), expanded the exponential in spherical harmonics and spherical Bessel functions, and examined the first term in the resulting infinite series. To calculate it, he extended the range of integration from V_s to infinity (equivalent to taking the limit $R_s \to \infty$), and found this term to be proportional to $(k_0\xi)^3 \ln (k_0\xi)$; all the other terms in the series were argued to have the same ξ dependence. However, since we have shown that the integral in (7.78) is proportional to R_s [compare (7.84)], Mountain's infinite series appears to be divergent. Also we find the sum of these terms depends on the temperature and wavelength through $(k_0\xi)^2$; finally we find that the $g^{(4)}$ term (which Mountain did not study) is larger and varies as $(k_0\xi)$.[4]

Recall that we assumed that $u^{(3)}$ and $u^{(4)}$ do not make significant long-range contributions to the depolarized intensity. Frisch and

McKenna[62] and Mountain[64] have suggested the possibility of using the depolarized intensity to study the behavior of these higher-distribution functions near the critical point. Our results, however, indicate that this procedure is not feasible. If we suppose for the moment that the Kirkwood superposition approximation[78] is valid near the critical point,

$$g^{(3)}(123) \simeq g(12)g(13)g(23), \tag{7.95}$$

then we have

$$u^{(3)}(123) \simeq u(12)u(23) + u(12)u(13) + u(13)u(23) + u(12)u(23)u(13) \tag{7.96}$$

A corresponding approximate equation can be written for $u^{(4)}$. If we insert these approximations into (7.70) it is easy to show that the $u^{(3)}$ and $u^{(4)}$ contributions do not involve volume dependence or delta functions and are smaller by a factor of $\sim 10^7$ R_s. Thus, unless $u^{(3)}$ and $u^{(4)}$ depart *significantly* from the superposition, approximation (*and* in fact have ranges considerably larger than a correlation length), their effect on the depolarized intensity will be completely masked by contributions from "iterated" $u^{(2)}$ terms.

We wish to stress in summary that the integrated light scattering intensity will in general contain three types of contributions through fourth order in the atomic polarizability: (*1*) α^2 terms corresponding to single scattering off undistorted particles—these terms commonly dominate the *polarized* spectrum, and vanish identically in the depolarized case; (*2*) α^4 terms arising from single scatterings off interacting clusters; and (*3*) α^4 terms corresponding to successive (double) scatterings off isolated particles separated by macroscopic distances. We found above that the type (*3*) contributions can easily dominate the depolarized spectrum of simple fluids near their critical points if the effective sample volume is large enough. It is important to stress, however, that if we were to rotate the polarizer at the detector by 90° we could find that I_R^{ZZ} [namely, the type (*1*) contribution] is in turn many orders of magnitude larger. Equivalently, in more visual terms, if the sample under conditions such that $I_R^{ZX}\,|(3) > I_R^{ZX}|\,(2)$, as discussed in detail above, were observed with the naked eye (which collects light of all polarizations) we would still see a neatly collimated laser beam rather than a glowing cell. Even though the *de*polarized intensity might be dominated by double scattering rather than collision-induced contributions, the *single* scattering off undistorted particles is still many time more intense. Thus, keeping only the lowest order nonvanishing terms [namely, α^4] in the expansion in powers of α is entirely valid.

It is obvious that the theory of multiple *polarized* scattering [i.e., where $I_R^{ZZ}\,|\,(1)$ no longer greatly exceeds $I_R^{ZZ}\,|\,(3)$] is in many ways parallel

to the theory that we have just presented for the depolarized intensity. The terms proportional to α^4 will be similar to (7.70), with the XZ-component of the dipole propagators replaced by the ZZ-component. [There will also be "cross-terms" between single and double scatterings, proportional to α^3.] By extending[73] our theory to the polarized case it should be possible to treat the "volume dependence" which has so often plagued experimental work near the critical point.[2] Higher-order scattering terms in α^5 and α^6 that become important for large enough R_s and small enough $(T_c - T)/T_c$ should also be calculable. Here, however, because of the reasons suggested in the above paragraph [namely, the slower convergence of the α-expansion] one has to be more careful in drawing quantitative conclusions.

D. Summary and Conclusions

In these final remarks we briefly restate the three principal points of view that we have attempted to develop in this article.

First, we have shown that all of the many, competing theories reviewed in Section III can be best understood and appreciated as special cases of general classical (see Section V) and/or quantum mechanical (Section IV) descriptions of depolarized light scattering by simple fluids. It is seen that all of the correlated-binary-collision-induced effects discussed in the literature follow directly from a two-body additive form for the adiabatic polarizability whose square defines the observed integrated intensity in the quantum mechanical theory. Similarly, the *classical* formulation of light propagation in simple fluids is used to demonstrate the equivalence of the effective (fluctuating) local field approach and the "multiple scattering" description. These latter theories are in turn shown to be identical to the long-range dispersional limit of the collision-induced models.

Second, we have attempted to describe and analyze as critically as possible (see Section VI) the various experimental data that have been used as probes of many-body susceptibilities and structure and/or dynamics in simple fluids. Unlike the first-order, polarized scattering effects that involve only the equilibrium correlations in a nontrivial way, the higher-order, *de*polarization features reflect the collective electromagnetic response as well. The ambiguities associated with simultaneously unraveling *both* the many-body polarizabilities *and* collisional events are considerable, and a great deal of careful analysis remains to be done. In fact, from the viewpoint of *new* information about simple fluids that has actually been gained from depolarization studies, very little of the "original promise" has been fulfilled. Nevertheless, certain *qualitative*

aspects of short-range polarizability distortion and many-body (radial/angular) correlations have been established as discussed in Section VI.

Third, in Section VII we have introduced in as detailed a manner as is presently possible several other phenomena that are intimately related to the collision-induced depolarization studies and thus help shed light on the general problems outlined above. In particular, we treated the density-dependent Kerr and dielectric functions as probes of pair *and* triplet, short-range polarizability distortion, *de*coupled from the complicating effects of many-body order. Similarly we considered the use of depolarized light scattering near the gas-liquid critical point as a test of long-range correlations involving *non*overlapping particles. The comprehensive study of nonlinear optical effects such as second harmonic generation and intensity-dependent refractive indices, and so on, should complement the above-mentioned analyses, but are of course subject to the same practical difficulties.

Since the completion of this review, two short papers have appeared which suggest approaches similar to ours for the unified treatment of nonresonant light scattering by many-body systems. Pasmanter, Samson, and Ben-Reuven[124] start with the usual second-order transition amplitude (much as we did in Section IV) for arbitrary polarizations and propagation vectors and rewrite it as the off-diagonal matrix element—between initial and final matter states—of a "macropolarizability" operator (A) which represents the linear electric susceptibility of the whole scattering sample treated as a single complex particle. Squaring and ensemble averaging (see Section IV.A) allows them to display the corresponding band shape as the Fourier transform of the autocorrelation function $\langle A^{+}(0)A(t)\rangle$. Upon ignoring intermolecular forces the operator A can be written as a sum of independent-particle contributions and the scattering cross-sections reduce trivially to N times the single molecule result; the local field corrections treated by Kirkwood[83] and Hellwarth[15] are then derived by including in A (or, more precisely, in the Liouville operator L_{sample} in terms of which A is defined) the dipole-dipole interaction responsible for excitation transfer among the molecules. It should be clear from the above remarks and the analogies with our earlier discussion that the "macropolarizability" approach of Pasmanter et al. can provide a unified description of *collision-induced* light scattering processes and other optical phenomena in which the sample's susceptibility can not be written as a sum of independent-particle (i.e., one-body) contributions. A rather different, but similarly general, treatment of light scattering by many-body systems has most recently been presented by Harris.[125] Here the calculation begins again with a transition amplitude (gotten by a (Van Vleck) transformation which brings in the matter-radiation interaction through terms

of second order) involving an adiabatic quantity very much like (4.39), say, but with our frequency-independent $\sum_{e \neq g}$ summation replaced by

$$\chi_{\alpha\beta}(\omega) \sim \sum_{e \neq g} \left\{ \frac{\langle g| \mathbf{x}_\alpha |e\rangle\langle e| \mathbf{x}_\beta |g\rangle}{(E_e - E_g) - \hbar\omega - i\delta} + \frac{\langle g| \mathbf{x}_\alpha |e\rangle\langle e| \mathbf{x}_\beta |g\rangle}{(E_e - E_g) + \hbar\omega} \right\} \quad (7.97)$$

with $\langle g|$ and $\langle e|$ still denoting the exact ground and excited electronic states for the *whole* system. In order to generalize upon the dipole limit, a density-density susceptibility, $\chi_{\alpha\beta}(\mathbf{r}\mathbf{r}'; \omega)$ is introduced which is defined by (7.97) but with $\mathbf{x}_{\alpha(\beta)} \to \rho_{\alpha(\beta)}[\mathbf{r}(\mathbf{r}')]$, the charge density operator associated with the αth (βth) particle center:

$$\chi_{\alpha\beta}(\omega) = \int d\mathbf{r} \int d\mathbf{r}' \mathbf{r}\mathbf{r}' \chi_{\alpha\beta}(\mathbf{r}\mathbf{r}'; \omega) \cdot \chi_{\alpha\beta}(\mathbf{r}\mathbf{r}'; \omega)$$

is then constrained to satisfy the time-dependent Hartree (random phase approximation) equation which is solved by iteration—the resulting expansion for the transition amplitude is shown to have the (now) familiar multiple scattering structure. The particular advantages of this Hamiltonian formulation derive from the (conceptual) ease with which one can include retardation effects, and optical processes involving two or more incident light beams.

Acknowledgments

We thank David Oxtoby and Donald Heller for their contributions to the work described in this chapter and, in particular, to Robert Harris for providing a consistently challenging sounding board for our ideas on the optical properties of simple fluids. George Birnbaum offered detailed comments on the original manuscript that were very helpful, as were many friendly arguments with Berni Alder and Herbert Strauss and several early discussions with Bruce Berne and John McTague. We gratefully acknowledge support from the Research Corporation and from the Donors of the Petroleum Research Fund, administered by the American Chemical Society.

APPENDIX A

We first rewrite (5.7) in the following way. The product of $\rho(\mathbf{r}'\tau)$ and $\mathbf{E}_0(\mathbf{r}'\tau)$ in the integrand [$\tau = t - (|\mathbf{R} - \mathbf{r}'/c)$, the retarded time] is trivially replaced by

$$\int_{-\infty}^{+\infty} dt' \rho(\mathbf{r}'t') \mathbf{E}_0(\mathbf{r}'t') \, \delta(t - \tau)$$

Then, introducing the Fourier integral representation of the delta function and noting that

$$\nabla_{\mathbf{R}} \times \nabla_{\mathbf{R}} \times \left[\frac{e^{ik|\mathbf{R}-\mathbf{r}'|}}{|\mathbf{R} - \mathbf{r}'|} \mathbf{v} \right] = \left[(\nabla_{\mathbf{R}}\nabla_{\mathbf{R}} + k^2 \mathbf{I}) \frac{e^{ik|\mathbf{R}-\mathbf{r}'|}}{|\mathbf{R} - \mathbf{r}'|} \right] \cdot \mathbf{v} \equiv [\mathsf{T}_k(\mathbf{R} - \mathbf{r}')] \cdot \mathbf{v} \quad (A.1)$$

where $\mathbf{v}$ is an arbitrary vector independent of $\mathbf{R}$, we have that (5.7) can be written as

$$\mathbf{E}(\mathbf{R}t) = \alpha \int_V d\mathbf{r}' \int_{-\infty}^{+\infty} dt' \mathsf{T}(\mathbf{R} - \mathbf{r}', t - t') \cdot \rho(\mathbf{r}'t')\mathbf{E}_0(\mathbf{r}'t') \qquad (A.2)$$

where

$$\mathsf{T}(\mathbf{R} - \mathbf{r}', t - t') \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} d(ck) e^{-ick(t-t')} \mathsf{T}_k(\mathbf{R} - \mathbf{r}') \qquad (A.3)$$

For $|\mathbf{R}| \gg |\mathbf{r}'|$ all $|\mathbf{r}'| \in V$ and the right-angle scattering geometry described in Fig. 1,

$$\mathsf{T}_k(\mathbf{R} - \mathbf{r}') \sim k^2 \frac{e^{ik(|\mathbf{R}| - \mathbf{r}' \cdot \hat{R})}}{|\mathbf{R}|} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (A.4)$$

The dyadic $(\mathsf{I} - \hat{R}\hat{R})$ on the right-hand side of (A.4) simply preserves the z-component of the linearly polarized incident light. Thus the only non-vanishing component of the first-order scattered field at the detector can be written $(k \to \bar{\omega}/c)$

$$E_z(\mathbf{R}t) = \frac{\alpha E_0}{2\pi c^2 R} \int_V d\mathbf{r}' \int_{-\infty}^{+\infty} dt' \int_{-\infty}^{+\infty} d\bar{\omega} \bar{\omega}^2 \times e^{i\mathbf{k} \cdot \mathbf{r}'} e^{-i\omega t} e^{i\bar{\omega}[t' - t + (R/c) - (\mathbf{r}' \cdot \mathbf{R}/cR)]} \rho(\mathbf{r}'t') \qquad (A.5)$$

Substituting (A.5) for E_z in (I) and performing the trivial integration over t gives

$$I_{\mathbf{R}}^{zz}(\omega') = \frac{\alpha^2 |\mathbf{E}_0|^2}{(2\pi R)^2} \frac{\omega^2}{c^4} \bar{\rho} \int_V d\mathbf{r}' \int_{-\infty}^{+\infty} dt' \int_V d\mathbf{r}'' \int_{-\infty}^{+\infty} dt'' \int_{-\infty}^{+\infty} d\omega'' \omega''^2 \times e^{i\mathbf{k} \cdot (\mathbf{r}' - \mathbf{r}'')} e^{-i\omega[t' - t'']} e^{i\omega'[t' + (R/c) - (\mathbf{R} \cdot \mathbf{r}'/Rc)]} \times e^{-i\omega''[t'' + (R/c) - (\mathbf{R} \cdot \mathbf{r}''/Rc)]} G^{VH}(|\mathbf{r}' - \mathbf{r}''|, t' - t'') \qquad (A.6)$$

Integrating over ω'' gives a factor $-(2\pi)(d^2/dt''^2)\ \delta[t'' + (R/C) - (\mathbf{R} \cdot \mathbf{r}''/Rc)]$, whereas integrating over the $\mathbf{r}'$- and t'-dependent functions gives

$$\exp(-i\mathbf{k}' \cdot \mathbf{r}'' + i\omega' t'') S(\mathbf{K}, \Omega)$$

where S is defined by (5.9). Noting that the ω'th Fourier component of $(d^2/dt'^2)\ \delta(t' - \tau)$ is $-\omega'^2 \exp(-i\omega'\tau)$, the remaining integrations over t'' and $\mathbf{r}''$ can be straightforwardly performed to give (5.8).

APPENDIX B

Using the asymptotic representation of $\mathsf{T}_k(\mathbf{R} - \mathbf{r}')$ (see Appendix A) and performing the integration over ck, $\hat{x} \cdot \mathsf{T}(\mathbf{R} - \mathbf{r}', t - t')\rho(\mathbf{r}'t')$ in

(5.12) can be written as

$$-\frac{1}{Rc^2}\frac{d^2}{dt'^2}\,\delta(t'-\tau)\rho(\mathbf{r}'t')P_x \tag{A.7}$$

where $\tau = t - (R/C) + (\hat{R}\cdot\mathbf{r}'/c)$. The operator P_x projects out the x-component of any vector to its right; P_x acting on the rest of the integrand can be expressed as

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} d(c\bar{k})\left\{\nabla_{\mathbf{r}'}\nabla_{\mathbf{r}'}\frac{e^{i\bar{k}|\mathbf{r}'-\mathbf{r}''|}}{|\mathbf{r}'-\mathbf{r}''|}\right\}_{xz} e^{-ic\bar{k}(t'-t'')}\rho(\mathbf{r}''t'')E_0(\mathbf{r}''t'') \tag{A.8}$$

But it is a simple exercise to show that the xz-component of the dyadic in curly brackets can be written as

$$[\mathsf{T}_{\bar{k}}(\mathbf{r}'-\mathbf{r}'')]_{xz} = \left\{\frac{\exp(i\bar{k}|\,\mathbf{r}'-\mathbf{r}''|)3(\mathbf{r}'-\mathbf{r}'')_x(\mathbf{r}'-\mathbf{r}'')_z}{|\mathbf{r}'-\mathbf{r}''|^5}\right\} \times\left\{1-i\bar{k}\,|\mathbf{r}'-\mathbf{r}''| - \frac{\bar{k}^2|\mathbf{r}'-\mathbf{r}''|^2}{3}\right\} \tag{A.9}$$

Substituting into (A.8) and performing the $c\bar{k}$ integration we pick up the factors of $2\pi\delta[t''-t'+(|\mathbf{r}'-\mathbf{r}''|/c)]$, $-(2\pi i/c)(d/dt'')\,\delta[t''-t'+(|\mathbf{r}'-\mathbf{r}''|/c)]$, and $-(2\pi/c^2)(d^2/dt''^2)\,\delta[t''-t'+(|\mathbf{r}'-\mathbf{r}''|/c)]$, respectively. We then integrate each of the three terms in the integrand over t''. Now, since the only other t''-dependent factor is $\rho(\mathbf{r}''t'')\mathbf{E}_0(\mathbf{r}''t'')$ and since the time variation of the fluctuations is small compared with that of the incident field ($\omega^{\text{optical}} \approx 10^{+15}\,\text{sec}^{-1}$) each integration involving $(d^{(n)}/dt''^{(n)})\,\delta(t''-t+|\mathbf{r}'-\mathbf{r}''|/c)$ gives a factor

$$(-i\omega)^n\rho(\mathbf{r}''t'')E_0(\mathbf{r}''t'')|_{t''=t'-|\mathbf{r}'-\mathbf{r}''|/c}$$

(recall that $\mathbf{E}_0(\mathbf{r}''t'') = \hat{z}E_0(\mathbf{r}''t'') = \hat{z}E_0\exp(i\mathbf{k}\cdot\mathbf{r}'')\exp(-i\omega t'')$). The condition that the variation in time of ρ be slow in comparison with $\mathbf{E}_0$ is almost always satisfied in light scattering problems: this restriction has in fact been implicitly assumed in writing down the integral form (5.1) of Maxwell's equations—see Born and Wolf. The right-hand side of (5.12) then becomes

$$-\frac{\alpha^2E_0}{Rc^2}\int_V d\mathbf{r}'\int_V d\mathbf{r}''\int_{-\infty}^{+\infty} dt'\,\frac{d^2}{dt'^2}\,\delta(t'-\tau)\rho(\mathbf{r}'t')[\mathsf{T}_{\omega/c}(\mathbf{r}'-\mathbf{r}'')]_{xz} \times\rho\left[\mathbf{r}'',\,t'+\frac{|\mathbf{r}'-\mathbf{r}''|}{c}\right]e^{i\mathbf{k}\cdot\mathbf{r}''}e^{-i\omega t'} \tag{A.10}$$

Performing the integration over t' and neglecting again changes in the

instantaneous density over times of the order of 10^{-15} sec, we have

$$\hat{x} \cdot \mathbf{E}(\mathbf{R}t) = \frac{\alpha^2}{R^2}\left(\frac{\omega}{c}\right)^2 \int_V d\mathbf{r}' \rho\left(\mathbf{r}', t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right) \times \int_{V-\sigma_{\mathbf{r}'}} d\mathbf{r}'' \rho\left(\mathbf{r}'', t - \frac{R}{c} + \frac{\hat{R} \cdot \mathbf{r}'}{c}\right) \times [\mathsf{T}_{\omega/c}(\mathbf{r}' - \mathbf{r}'')]_{xz} E_0 e^{i\mathbf{k}\cdot\mathbf{r}''} e^{-i\omega[t-(R/c)+(\hat{R}\cdot\mathbf{r}'/c)]} + \theta(\alpha^3) \quad \text{(A.11)}$$

Writing the instantaneous one-particle densities as sums over the appropriate delta functions, (5.19) follows trivially.

APPENDIX C

Recall that [see (4.35)]

$$A_{xz}(\{\mathbf{R}^N\}) = \hat{x} \cdot \sum_{e \neq g} \frac{\langle g | \sum_i \mathbf{p}_i e^{-i\mathbf{k}'\cdot\mathbf{x}_i} | e \rangle \langle e | \sum_j \mathbf{p}_j e^{i\mathbf{k}\cdot\mathbf{x}_j} | g \rangle}{(E_e - E_g) - \hbar\omega} \cdot \hat{z} \quad \text{(A.12)}$$

where $\mathbf{p}_i$ and $\mathbf{x}_i$ denote the momentum and position of the ith electron in the fluid; i runs over all electrons. $\langle g|$ and $\{\langle e|\}$ are the adiabatic ground and excited electronic states of the system. Now, we want to derive the limiting form to which A_{xz} reduces in the case where only particles 1 and 2 are present and interacting (i.e., these two atoms are within, say, 10 Å of one another). Then we can write

$$\sum_i \mathbf{p}_i e^{-i\mathbf{k}'\cdot\mathbf{x}_i} \to e^{-i\mathbf{k}'\cdot\mathbf{R}_{12}} \sum_{i \in 1,2} \mathbf{p}_i e^{-i\mathbf{k}'\cdot(\mathbf{x}_i - \mathbf{R}_{12})} \approx e^{-i\mathbf{k}'\cdot\mathbf{R}_{12}} \mathbf{p}_{12} \quad \text{(A.13)}$$

Here $\mathbf{R}_{12}$ denotes a vector drawn from the origin to the center of mass of the interacting pair of atoms 1, 2; $\mathbf{p}_{12}$ is the total momenta of the electrons associated with the pair. We have simply used the fact that $\mathbf{k}' \cdot (\mathbf{x}_i - \mathbf{R}_{12}) \ll 1$ since $|\mathbf{k}'| \approx 10^5$ cm^{-1} for optical wavelengths while $|\mathbf{x}_i - \mathbf{R}_{12}| \leqslant 10^7$ cm for interacting atoms. Thus we can write (after simplifying $\sum_j \mathbf{p}_j e^{i\mathbf{k}\cdot\mathbf{x}_j}$ similarly)

$$A_{xz} = e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{12}} \hat{x} \cdot \sum_{e \neq g} \frac{\langle g^{12} | \mathbf{p}_{12} | e^{12} \rangle \langle e^{12} | \mathbf{p}_{12} | g^{12} \rangle}{(E_{e^{12}} - E_{g^{12}}) - \hbar\omega} \cdot \hat{z} \quad \text{(A.14)}$$

(The 1, 2 superscripts refer of course to the adiabatic separation of electronic and nuclear motions having been applied to the diatom 1, 2.)

Now, for $r_e \equiv \hbar\omega/(E_{e1,2} - E_{g1,2})$ sufficiently less than unity (e.g., $r_e \lessapprox \frac{1}{4}$ for the monatomic fluids and optical wavelengths) we have that

$$\frac{1}{(E_{e1,2} - E_{g1,2}) - \hbar\omega} \approx \frac{1}{(E_{e1,2} - E_{g1,2})} \times \left[1 + \frac{\hbar\omega}{(E_{e1,2} - E_{g1,2})} + \frac{\hbar^2\omega^2}{(E_{e1,2} - E_{g1,2})^2}\right] \quad \text{(A.15)}$$

But then

$$\hat{x} \cdot \sum_{e \neq g} \frac{\langle g^{1,2}| \mathbf{p}_{12} |e^{1,2}\rangle\langle e^{1,2}| \mathbf{p}_{12} |g^{1,2}\rangle}{(E_{e^{1,2}} - E_{g1,2}) - \hbar\omega} \cdot \hat{z} = T_1 + T_2 + T_3 \qquad (A.16)$$

where

$$T_1 = \sum_{e \neq g} \frac{\langle g^{1,2}| \hat{x} \cdot \mathbf{p}_{12} |e^{1,2}\rangle\langle e^{1,2}| \mathbf{p}_{12} \cdot \hat{z} |g^{1,2}\rangle}{(E_{e1,2} - E_{g1,2})}$$

$$= \frac{im}{\hbar} \hat{x} \cdot \langle g^{1,2}| \mathbf{px} |g^{1,2}\rangle \cdot \hat{z} \qquad (A.17)$$

$$T_2 = \hbar\omega \sum_{e \neq g} \frac{\langle g^{1,2}| \hat{x} \cdot \mathbf{p}_{12} |e^{1,2}\rangle\langle e^{1,2}| \mathbf{p}_{12} \cdot \hat{z} |g^{1,2}\rangle}{(E_{e1,2} - E_{g1,2})^2}$$

$$= \frac{m^2\omega}{\hbar} \hat{x} \cdot \langle g^{1,2}| \mathbf{xx} |g^{1,2}\rangle \cdot \hat{z} \equiv \frac{m^2\omega}{\hbar} Q_{xz}^{1,2} \qquad (A.18)$$

and

$$T_3 = \hbar^2\omega^2 \sum_{e \neq g} \frac{\langle g^{1,2}| \hat{x} \cdot \mathbf{p}_{12} |e^{1,2}\rangle\langle e^{1,2}| \mathbf{p}_{12} \cdot \hat{z} |g^{1,2}\rangle}{(E_{e1,2} - E_{g1,2})^3}$$

$$= \omega^2 m^2 \hat{x} \cdot \sum_{e \neq g} \frac{\langle g^{1,2}| \mathbf{x} |e^{1,2}\rangle\langle e^{1,2}| \mathbf{x} |g^{1,2}\rangle \cdot \hat{z}}{(E_{e1,2} - E_{g1,2})} \equiv \frac{\omega^2 m^2}{2e^2} \alpha_{xz}^{1,2} \qquad (A.19)$$

In the above we have simply invoked the well-known identity $\langle e| \mathbf{p} |g\rangle = (im/\hbar)(E_e - E_g)\langle e| \mathbf{x} |g\rangle$ where m is the electron mass. Furthermore, from the properties of the commutator $[\mathbf{p}, \mathbf{x}]$ and the assumption of real electronic wavefunctions (namely, $\langle e| \mathbf{x} |g\rangle$ real and $\langle g| \mathbf{p} |e\rangle$ pure imaginary, etc.) it follows that $T_1 = (m/2)\,\delta_{xz} = 0$. Thus we obtain [see (4.38)]

$$A_{xz} = e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{12}} \left[\frac{\omega}{2} \left(\frac{m}{e} \right)^2 \alpha_{xz}^{1,2} + \frac{\omega m^2}{\hbar} Q_{xz}^{1,2} \right] \qquad (A.20)$$

and

$$A_{xz}{}^{+} = (A_{xz})^* \qquad (A.21)$$

Everything follows through similarly for B_{zx} and $B_{zx}{}^{+}$, giving rise to (4.42).

APPENDIX D

We wish to calculate

$$g(\mathbf{k}_1) = \int_{V_s - \sigma_0} d\mathbf{r} e^{i\mathbf{k}_1 \cdot \mathbf{r}} [\mathsf{T}_{k_0}(\mathbf{r})]_{XZ} \qquad (A.22)$$

where the tensor $\mathsf{T}_{k_0}(\mathbf{r})$ is defined in (7.67) of the text.

It is convenient to choose a new coordinate system (x', y', z') such that $z' \parallel k_1$, and $y' \perp z$ (x, y, z still refer to the original scattering geometry shown in Fig. 1). Let $\mathbf{k}_1$ be defined by the polar angles (θ, ϕ) in the x, y, z

coordinate system. Then

$$\begin{aligned}\hat{z}' &= \hat{k}_1 = \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z}\\ \hat{y}' &= -\sin\phi\,\hat{x} + \cos\phi\,\hat{y}\\ \hat{z}' &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\hat{z};\end{aligned} \tag{A.23}$$

as $\theta, \phi \to 0$ the x', y', z') system reduces to the (x, y, z) system.

Equation (A.23) can be inverted to find $\hat{x}, \hat{y}, \hat{z}$ in terms of the primed coordinates. Writing

$$\mathbf{r} = \sin\theta'\cos\phi'\,\hat{x}' + \sin\theta'\sin\phi'\,\hat{y}' + \cos\theta'\,\hat{z}' \tag{A.24}$$

we can easily show that

$$\mathbf{k}_1 \cdot \mathbf{r} = k_1 r\cos\theta' \tag{A.25}$$

For the tensor quantity, we need to calculate r_x and r_z:

$$\begin{aligned}r_z &= \mathbf{r}\cdot\hat{z} = r(\cos\theta\cos\theta' - \sin\theta\sin\theta'\cos\phi')\\ r_x &= \mathbf{r}\cdot\hat{x} = r(\sin\theta\cos\phi\cos\theta' - \sin\phi\sin\theta'\sin\phi'\\ &\qquad + \cos\theta\cos\phi\sin\theta'\cos\phi')\end{aligned} \tag{A.26}$$

Integrating (A.22) over ϕ' gives

$$\begin{aligned}&\pi\int_0^{\pi} d\theta'\sin\theta'\int_{\sigma}^{Rs} drr^2 e^{ik_0 r}\left(\frac{3}{r^3} - \frac{3ik_0}{r^2} - \frac{{k_0}^2}{r}\right)\exp\left[ik_1 r\cos\theta'\right]\\ &\quad\times\left[2\sin\theta\cos\theta\cos\phi\cos^2\theta' - \sin\theta\cos\theta\cos\phi\sin^2\theta'\right]\\ &= 2\pi\sin\theta\cos\theta\cos\phi\int_{\sigma}^{Rs} dre^{ik_0 r}\left(\frac{3}{r} - 3ik_0 - {k_0}^2 r\right)\\ &\quad\times\left\{e^{ik_1 r}\left[\frac{1}{ik_1 r} - \frac{3}{(ik_1 r)^2} + \frac{3}{(ik_1 r)^3}\right]\right.\\ &\quad\left. + e^{-ik_1 r}\left[-\frac{1}{ik_1 r} - \frac{3}{(ik_1 r)^2} - \frac{3}{(ik_1 r)^3}\right]\right\}\\ &= 2\pi\sin\theta\cos\theta\cos\phi\int_{\sigma}^{Rs} dr\left\{\exp\left[i(k_0 + k_1)r\right]\right.\\ &\quad\times\left[\frac{1}{r^4}\left(\frac{9i}{{k_1}^3}\right) + \frac{1}{r^3}\left(\frac{9k_0}{{k_1}^3} + \frac{9}{{k_1}^2}\right) + \frac{1}{r^2}\left(-\frac{3i{k_0}^2}{{k_1}^3} - \frac{9ik_0}{{k_1}^2} - \frac{3i}{k_1}\right)\right.\\ &\quad\left.\left. + \frac{1}{r}\left(-\frac{3{k_0}^2}{{k_1}^2} - \frac{3k_0}{k_1}\right) + \frac{ik_0^2}{k_1}\right]\right.\end{aligned}$$

$$+ \exp[i(k_0 - k_1)r]\left[\frac{1}{r^4}\left(-\frac{9i}{{k_1}^3}\right) + \frac{1}{r^3}\left(-\frac{9k_0}{{k_1}^3} + \frac{9}{{k_1}^2}\right)\right.$$

$$\left. + \frac{1}{r^2}\left(\frac{3i{k_0}^2}{{k_1}^3} - \frac{9ik_0}{{k_1}^2} + \frac{3i}{k}\right) + \frac{1}{r}\left(-\frac{3{k_0}^2}{{k_1}^2} + \frac{3k_0}{k_1}\right) - \frac{i{k_0}^2}{k_1}\right]\Bigg\}$$

$$= 2\sin\theta\cos\theta\cos\phi\left[-6e^{iy_0}\frac{\sin y_1}{{y_1}^3} + 6e^{iy_0}\frac{\cos y_1}{{y_1}^2}\right.$$

$$- \frac{{y_0}^2}{y_1(y_0 + y_1)}\cos(y_0 + y_1) + \frac{{y_0}^2}{y_1(y_0 - y_1)}\cos(y_0 - y_1)\Bigg]$$

$$+ i\left[6y_0e^{iy_0}\frac{\sin y_1}{{y_1}^3} - 6y_0e^{iy_0}\frac{\cos y_1}{y_0^2}\right.$$

$$\left. - \frac{{y_0}^2}{y_1(y_0 + y_1)}\sin(y_0 + y_1) + \frac{{y_0}^2}{y_1(y_0 - y_1)}\sin(y_0 - y_1)\right]\Bigg\} \quad \text{(A.27)}$$

Here we introduced $y_0 = k_0\sigma \ll 1$, $y_1 = k_1\sigma$. Note that the last step is only valid for $k_1 \neq k_0$.

The above is a fairly well-behaved function. As $k_1 \to 0$, it approaches zero, while for $k_1 \to \infty$ it becomes

$$\sim 2\pi\sin\theta\cos\theta\cos\phi\left[\frac{6\cos(k_1\sigma)}{(k_1\sigma)^2}\right]$$

so that $g(\mathbf{k}_1)$ should be square integrable. The only singularity comes at $k_1 = k_0$, from the term

$$2\pi\sin\theta\cos\theta\cos\phi\left(\frac{-i{k_0}^2}{k_1}\right)\int_\sigma^{Rs} dr e^{i(k_0 - k_1)r}$$

If $k_0 \simeq k_1$, we must treat the upper limit exactly, giving

$$2\pi\sin\theta\cos\theta\cos\phi\left(\frac{-{k_0}^2}{k_1}\right)\frac{\exp[i(k_0 - k_1)R_s] - \exp[i(k_0 - k_1)\sigma]}{k_0 - k_1}$$

When $g(k_1)$ is squared, the result is

$$|g(\mathbf{k}_1)|^2 = \sin^2\theta\cos^2\theta\cos^2\phi$$

$$\times \Big\{4\pi^2\,[\text{nonsingular terms in (A.27)}]^2 + 8\pi^2\,(\text{cross terms})$$

$$+ 4\pi^2\left(\frac{{k_0}^4}{{k_1}^2}\right)\left|\int_\sigma^{Rs} dr \exp[i(k_0 - k_1)r]\right|^2\Big\}$$

$$= \sin^2\theta\cos^2\theta\cos^2\phi$$

$$\times \left[h(k_1) + 4\pi^2\frac{{k_0}^4}{{k_1}^2(k_0 - k_1)^2}\left|e^{i(k_0 - k_1)Rs} - e^{i(k_0 - k_1)\sigma}\right|^2\right. \quad \text{(A.28)}$$

As discussed above, $h(k_1)$ is well-behaved except for a delta function at k_0 arising from the cross terms.

The last part of (A.28) can be calculated explicitly: it is

$$4\pi^2\left(\frac{k_0^4}{k_1^2}\right)\frac{2 - 2\cos(k_0 - k_1)(R_s - \sigma)}{(k_0 - k_1)^2} \simeq 8\pi^2\left(\frac{k_0^2}{k_1^2}\right)\frac{1 - \cos(k_0 - k_1)R_s}{(k_0 - k_1)^2} \tag{A.29}$$

which is finite at $k_1 = k_0$. It has a series of peaks at

$$k_0 - k_1 = \frac{(2n + 1)\pi}{R_s}, \qquad n = 0, \pm 1, \pm 2, \ldots$$

whose heights are

$$\sim \frac{16k_0^2R_s^2}{(2n + 1)^2}$$

Since R_s is large, these peaks are very high and narrow: furthermore, since the function that $|g(k_1)|^2$ multiplies in (7.80) and (7.89) is a smooth function in the region $k_1 \simeq k_0$, the *effect* of these peaks is simply that of a delta function. We calculate the normalization factor C by requiring that

$$\int_{-\infty}^{\infty} dk_1 C\delta(k_1 - k_0) = C = \int_{-\infty}^{\infty} dk_1 8\pi^2 k_0^2 \frac{1 - \cos(k_0 - k_1)R_s}{(k_0 - k_1)^2}$$

$$= 8\pi^2 k_0^2 R_s \int_{-\infty}^{\infty} dy \frac{1 - \cos y}{y} = 8\pi^3 k_0^2 R_s \tag{A.30}$$

Thus

$$|g(\mathbf{k}_1)|^2 = \sin^2\theta \cos^2\theta \cos^2\phi[h(k_1) + 8\pi^3 k_0^2 R_s \delta(k_1 - k_0)] \tag{A.31}$$

which is equivalent to (7.82).

The dependence on the sample volume arises because $|g(\mathbf{k}_1)|^2$ contains the product of two functions closely resembling delta functions, which have the same arguments. A term such as $|g(\mathbf{k}_1)g^*(\mathbf{k}_2)|$ where $\mathbf{k}_1 \neq \mathbf{k}_2$ would not have this dependence on R_s, and it is for this reason that the second and third $g^{(3)}$ terms in (7.70) and the second contribution to the $g^{(4)}$ term (from $u(14)\,u(23)$) are negligible in comparison to the R_s-dependent terms. The fourth $g^{(3)}$ term in (7.70) is small because of the restriction that $\mathbf{r}_{12} + \mathbf{r}_{23}$ lie inside V_I, rather than taking on any value in V_s.

References

1. See, for example, I. L. Fabelinskii, *Molecular Scattering of Light*, Plenum, New York, 1968, and references cited in B. Chu, *Ann. Rev. Phys. Chem.*, **21**, 145 (1970). See also Ref. 25.
2. D. McIntyre and J. V. Sengers, in *Physics of Simple Liquids*, H. N. V. Temperley et al., Eds., North-Holland, Amsterdam, 1968; R. D. Mountain, Critical Reviews,

Solid State Sciences, **1** (1), 5 (1970); other pertinent references will be cited where appropriate throughout the text.
3. (*a*) J. P. McTague and G. Birnbaum, *Phys. Rev. Lett.*, **21**, 661 (1968); (*b*) *Phys. Rev.*, **A3**, 1376 (1971).
4. P. Lallemand, *Phys. Rev. Lett.*, **25**, 1079 (1970); *J. Phys. Radium*, **32**, 119 (1971).
5. J. I. Gersten, R. E. Slusher, and C. M. Surko, *Phys. Rev. Lett.*, **25**, 1739 (1970).
6. V. Volterra, J. A. Bucaro, and T. A. Litovitz, *Phys. Rev. Lett.*, **26**, 55 (1971).
7. (*a*) M. Thibeau, G. C. Tabisz, B. Oksengorn, and B. Vodar, *J. Quant. Spectr. Radiative Transfer*, **10**, 839 (1970); M. Thibeau, B. Oksengorn, and B. Vodar, *J. Phys. Radium*, **30**, 47 (1969); *ibid.*, **29**, 287 (1968). (*b*) M. Thibeau and B. Oksengorn, *Mol. Phys.*, **15**, 579 (1968).
8. J. P. McTague, W. D. Ellenson, and L. H. Hall, *J. Phys., Radium*, **33**, Colloque C-1, 241 (1972).
9. J. P. McTague, P. A. Fleury, and D. B. DuPre, *Phys. Rev.*, **188**, 303 (1969).
10. W. S. Gornall, H. E. Howard-Lock, and B. P. Stoicheff, *Phys. Rev.*, **A1**, 1288 (1970).
11. J. A. Bucaro and T. A. Litovitz, *J. Chem. Phys.*, **54**, 3846 (1971).
12. P. A. Fleury, W. B. Daniels, and J. M. Worlock, *Phys. Rev. Lett.*, **27**, 1493 (1971).
13. G. C. Tabisz, W. R. Wall, D. P. Shelton, and J. H. K. Ho, "The Spectral Profile of Collision-Induced Light Scattering from Some Molecular Liquids," paper presented at Third International Conference on Raman Spectroscopy, Reims, France, September, 1972. [We are grateful to Prof. Tabisz for a preprint copy.]
14. (*a*) H. B. Levine and G. Birnbaum, *Phys. Rev. Lett.*, **20**, 439 (1968); (*b*) *J. Chem. Phys.*, **55**, 2914 (1971). See also Ref. 3.
15. R. W. Hellwarth, *J. Chem. Phys.*, **52**, 2128 (1970).
16. (*a*) J. C. Lewis and J. van Kranendonk, *Phys. Rev. Letters* **24**, 802 (1970). (b) Evidence for the small frequency "dip" predicted by Lewis and van Kranendonk in their discussion of collisional interference effects in the depolarized line shape has been reported recently by D. J. G. Kerwin and A. D. May [*Can. J. Phys.* **50**, 2174 (1972)]—we are grateful to Dr. George Birnbaum for pointing out this latter reference to us.
17. C. G. Gray and H. I. Ralph, *Phys. Lett.*, **33A**, 165 (1970).
18. P. Lallemand, *Compt. Rend.*, **273**, 89 (1971); *J. Phys. Radium*, **33**, Colloque C-1, 257 (1972). See also Ref. 4.
19. J. I. Gersten, *Phys. Rev.*, **A4**, 98 (1971).
20. H. K. Shin, *J. Chem. Phys.*, **56**, 2617 (1972).
21. W. M. Gelbart, *J. Chem. Phys.*, **57**, 699 (1972).
22. B. J. Alder, J. J. Weis, and H. L. Strauss, *Phys. Rev.*, **A7**, 281 (1973).
23. Background work and other related papers will be referenced where appropriate throughout the text.
24. A. Einstein, *Ann. Physik*, **33**, 1275 (1910).
25. G. B. Benedek, in *Statistical Physics, Phase Transitions and Superfluidity*, Vol. 2, M. Chrétien et al., Eds., Gordon and Breach, New York, 1968.
26. See, for example, J. Cabannes, *La Diffusion Molèculaire de la Lumière*, Les Presses Universitaires de France, 1929.
27. R. D. Mountain, *Rev. Mod. Phys.*, **38**, 205 (1966).
28. L. P. Kadanoff and P. C. Martin, *Ann. Phys. N.Y.*, **24**, 419 (1963).
29. S. Yip and M. Nelkin, *Phys. Rev.*, **135**, A1241 (1964); M. Nelkin and S. Ranganathan, *Phys. Rev.*, **164**, 222 (1967) and *J. Chem. Phys.*, **47**, 4056 (1967); M. Nelkin and P. J. Ortoleva, *IAEA Symposium on Inelastic Neutron Scattering*

(IAEA, Copenhagen, 1969), International Atomic Energy Agency, Vienna, Austria, and references contained therein.
30. (*a*) R. Pecora and W. A. Steel, *J. Chem. Phys.*, **42,** 1872 (1965). (*b*) See also R. G. Gordon, *J. Chem. Phys.*, **42,** 3658 (1965) in the case where his molecule stays in its ground vibrational state.
31. S. L. Shapiro and H. P. Broida, *Phys. Rev.*, **154,** 129 (1967).
32. D. A. Pinnow, S. J. Candau, and T. A. Litovitz, *J. Chem. Phys.*, **49,** 347 (1968).
33. A. Ben-Reuven and N. D. Gershon, *J. Chem. Phys.*, **51,** 893 (1969).
34. See also the extensive Russian literature on rotational Brownian motion and anisotropy fluctuations referred to in, for example, V. S. Starunov, *Soviet Phys. "Doklady" English Transl.*, **8,** 1206 (1964).
35. V. S. Starunov, E. V. Tiganov, and I. L. Fabelinskii, *JETP Lett.*, **4,** 176 (1966) [English transl.].
36. A. Szöke, E. Courtens, and A. Ben-Reuven, *Chem. Phys. Lett.*, **1,** 87 (1967).
37. G. I. A. Stegeman and B. P. Stoicheff, *Phys. Rev. Lett.*, **21,** 202 (1968).
38. A. Ben-Reuven and N. D. Gershon, *J. Chem. Phys.*, **55,** 475 (1971) *ibid.*, **53,** 3397 (1971). See also Ref. 33.
39. V. Volterra, *Phys. Rev.*, **180,** 156 (1969); H. C. Andersen and R. Pecora, *J. Chem. Phys.*, **54,** 2584 (1971), C. H. Chung and S. Yip, *Phys. Rev.*, **A4,** 928 (1971); and T. Keyes and D. Kivelson, *J. Chem. Phys.*, **54,** 1786 (1971). The latter discussions treat the shear mode as dissipative rather than propagating.
40. N. K. Ailawadi, B. J. Berne, and D. Forster, *Phys. Rev.*, **A3,** 1472 (1971).
41. Proceedings of the International Colloquim of the C.N.R.S. on The Scattering of Light by Fluids (July 15–17, 1971, Paris), published in *J. Phys. Radium*, **33,** C-1 (1972). See in particular the session on "Depolarized Rayleigh Scattering," pp. Cl-181 through Cl-240.
42. For a recent discussion see, for example, *Raman Spectroscopy*, Vol. 2, H. A. Szymanski, Ed., Plenum Press, New York, 1970.
43. B. J. Berne and H. L. Frisch, *J. Chem. Phys.*, **47,** 3675 (1967); B. J. Berne, J. M. Deutch, J. T. Hynes, and H. L. Frisch, *J. Chem. Phys.*, **49,** 2864 (1968); B. J. Berne and R. Pecora, *J. Chem. Phys.*, **50,** 783 (1969); *ibid.*, **51,** 475 (1969).
44. J. M. Schurr, *J. Phys. Chem.*, **73,** 2820 (1969).
45. L. Blum and Z. W. Salsburg, *J. Chem. Phys.*, **48,** 2292 (1968); *ibid.*, **50,** 1654 (1969); L. Blum, *J. Chem. Phys.*, **51,** 5024 (1969); and D. L. Knirk and Z. W. Salsburg, *J. Chem. Phys.*, **54,** 1251 (1971).
46. M. Weinberg and R. Kapral, *J. Chem. Phys.*, **53,** 4409 (1970); R. Kapral, *J. Chem. Phys.*, **56,** 1842 (1972).
47. A. G. Marshall and R. Pecora, *J. Chem. Phys.*, **55,** 1245 (1971). For a more general discussion of the light scattering properties of flexible large molecules see, for example, R. Pecora, *J. Chem. Phys.*, **49,** 1032 (1968), *Macromol.*, **2,** 31 (1969); and M. J. Stephen, *J. Chem. Phys.*, **55,** 3585 (1971), and references to related theoretical and experimental work cited therein.
48. P. A. Franken and J. F. Ward, *Rev. Mod. Phys.*, **35,** 23 (1963).
49. W. L. Peticolas, *Ann. Rev. Phys. Chem.*, **18,** 233 (1967).
50. N. Bloembergen, *Nonlinear Optics*, Benjamin, New York, 1965.
51. W. M. Gelbart, *Chem. Phys. Letters* **23,** 53 (1973).
52. M. F. Crawford, H. L. Welsh, and J. L. Locke, *Phys. Rev.*, **75,** 1607 (1949).
53. For more recent papers containing references to earlier work see E. J. Allin, A. D. May, B. P. Stoicheff, J. L. Stryland, and H. L. Welsh, *Appl. Optics*, **6,** 1597 (1967);

A. Rosenberg and G. Birnbaum, *J. Chem. Phys.*, **48,** 1396 (1968); and H. A. Gebbie and N. W. B. Stone, *Proc. Phys. Soc., London*, **82,** 543 (1963).
54. See, for example, work cited in H. Levine, *J. Chem. Phys.*, **56,** 2455 (1972) and *ibid* **55,** 1039 (1971).
55. G. Birnbaum, H. Levine, and D. A. McQuarrie, *J. Chem. Phys.*, **46,** 1557 (1967).
56. M. Fixman, *J. Chem. Phys.*, **23,** 2074 (1955): contains reference to earlier work.
57. A. D. Buckingham and M. J. Stephen, *Trans. Faraday Soc.*, **53,** 884 (1957).
58. S. Kielich, *J. Chem. Phys.*, **46,** 4090 (1967); *Acta Phys. Polon.*, **19,** 149 (1960).
59. O. Theimer and R. Paul, *J. Chem. Phys.*, **42,** 2508 (1965).
60. P. Mazur, *Advances in Chemical Physics.*, Vol. I, I. Prigogine and J. A. Rice, Eds., Interscience, New York, 1958, p. 309.
61. M. Tanaka, *Progr. Theor. Phys.*, **40,** 975 (1968).
62. H. L. Frisch and J. McKenna, *Phys. Rev.*, **139,** A68 (1965).
63. R. K. Bullough, *Phil. Trans. Roy. Soc. London Ser.* **A 258,** 387 (1965), and references cited therein.
64. R. Mountain, *J. Phys. Radium*, **33,** Colloque C-1, 265 (1972).
65. M. J. Stephen, *Phys. Rev.*, **187,** 279 (1969).
66. See H. B. Levine, *Phys. Rev.*, **160,** 159 (1967); Refs. 52–55.
67. A. D. Buckingham and D. A. Dunmur, *Trans. Faraday Soc.*, **64,** 1776 (1968).
68. J. Van Kranendonk, *Can. J. Phys.*, **46,** 1173 (1968); see also V. F. Sears, *Can. J. Phys.*, p. 1163.
69. V. Volterra, J. A. Bucaro, and T. A. Litovitz, *Ber. Bunsenges. Physik. Chem.*, **75,** 309 (1971).
70. B. Berne, M. Bishop and A. Rahman, *J. Chem. Phys.* **58,** 2696 (1973).
71. See, for example, J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading, Mass., 1967, Chap. 2.
72. The standard reference is M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, London, 1954.
73. D. Oxtoby and W. M. Gelbart, unpublished results.
74. L. Van Hove, *Phys. Rev.*, **95,** 249 (1954).
75. L. I. Komarov and I. Z. Fisher, *Soviet Phys. JETP English Transl.*, **16,** 1358 (1963).
76. R. Pecora, *J. Chem. Phys.*, **40,** 1604 (1964).
77. See, for example, the reviews by H. H. Paalman and C. J. Pings, *Rev. Mod. Phys.*, **35** (2), 389 (1963).
78. See, for example, T. L. Hill, *Statistical Mechanics*, McGraw-Hill, New York, 1956.
79. W. M. Gelbart, *Mol. Phys.*, **26,** 873 (1973).
80. E. F. O'Brien, V. P. Gutschick, V. McKoy, and J. P. McTague, *Phys. Rev.* **A8,** 690 (1973).
81. S. Kielich, *Acta Phys. Polon.*, **32,** 297 (1967); *IEEE J. Quantum Electron.*, **QE-4,** 744 (1968).
82. J. D. Ramshaw, *J. Chem. Phys.*, **57,** 2070 (1972). It should be noted that our cluster analysis is similar to that employed by J. D. Poll and J. van Kranendonk, Can. J. Phys., **39,** 189 (1961), in calculating the dipole moment which gives rise to the pressure-dependent, collision-induced absorption coefficients of gases. See also Ref. 8.
83. J. G. Kirkwood, *J. Chem. Phys.*, **4,** 592 (1936) and Kirkwood Collected Works, Vol. 2, *Dielectrics: Intermolecular Forces—Optical Rotation*, R. H. Cole, Ed., Gordon and Breach, New York, 1965.

84. J. Yvon, *Actualités Scientifique et Industrielles*, Nos. 652 and 543, Hermann, Paris, 1937. See also discussion by W. F. Brown, Jr., *J. Chem. Phys.*, **18,** 1193 (1950); and H. S. Green, *The Molecular Theory of Fluids*, Interscience, New York, 1952, pp. 207–213.
85. L. Rosenfeld, *Theory of Electrons*, North-Holland, Amsterdam, 1951.
86. M. Born and E. Wolf, *Principle of Optics*, Pergamon, New York, 1964.
87. C. G. Darwin, *Trans. Cambridge Phil. Soc.*, **23,** 137 (1924).
88. L. Jansen and P. Mazur, *Physica*, **21,** 193, 208 (1955); P. Mazur and M. Mandel, *Physica*, **22,** 289, 299 (1956).
89. G. H. Vineyard, *Phys. Rev.*, **96,** 93 (1954). See also discussion by J. E. Enderby in *Physics of Simple Liquids*, H. N. V. Temperley et al., Eds., North-Holland, Amsterdam, 1968; and R. J. Glauber in *Boulder Lectures in Theoretical Physics*, Vol. 4, W. E. Brittin, Ed., Interscience, New York, 1962.
90. (*a*) K. Sköld and K. E. Larsson, *Phys. Rev.*, **161,** 102 (1967).
90. (*b*) K. Sköld, J. M. Rowe, G. Ostrowski, and P. D. Randolph, *Phys. Rev.*, **A6** (3), 1107 (1972).
91. M. A. Gray, T. M. Loehr and P. A. Pincus, *J. Chem. Phys.* **59,** 1121 (1973).
92. (a) G. Birnbaum (private communication). (b) G. Birnbaum and E. R. Cohen, preprint entitled "Models for the Anisotropic Polarizability Induced in Colliding Pairs of Rare Gas Atoms".
93. J. K. Percus and G. J. Yevick, *Phys. Rev.*, **110,** 1 (1958).
94. P. G. deGennes, *Physica*, **25,** 825 (1959).
95. P. G. Mikolaj and C. J. Pings, *J. Chem. Phys.*, **46,** 1401 (1967); N. S. Gingrich and C. W. Thompson, *J. Chem. Phys.*, **36,** 2398 (1962).
96. See, for example, B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, **31,** 459 (1959).
97. B. J. Alder, H. L. Strauss, and J. J. Weis, *J. Chem. Phys.* **59,** 1002 (1973).
98. T. Keyes, D. Kivelson, and J. P. McTague, *J. Chem. Phys.*, **55,** 4096 (1971).
99. H. Mori, *Progr. Theoret. Phys.*, **33,** 423 (1965). See also the exposition of this work by B. J. Berne in *Physical Chemistry: An Advanced Treatise*, H. Eyring et al., Eds., Academic Press, New York, 1971.
100. See, for example, W. A. Steele, *J. Chem. Chem.*, **39,** 3197 (1963).
101. J. Kerr, *Phil. Mag.*, **50,** 337, 446 (1875).
102. M. Born, *Ann. Phys.* (*Leipzig*), **55,** 177 (1918); P. Langevin, *J. Radium, Paris*, **7,** 249 (1910).
103. A. D. Buckingham and J. A. Pople, *Proc. Phys. Soc.*, **A68,** 905 (1955).
104. A. D. Buckingham, *Proc. Phys. Soc.*, **A68,** 910 (1955).
105. P. Mazur and B. J. Postma, *Physica*, **25,** 251 (1959).
106. See R. H. Orcutt and R. H. Cole, *J. Chem. Phys.*, **46,** 697 (1967) and references to earlier work cited therein.
107. D. Henderson and L. Oden, *Mol. Phys.*, **10,** 405 (1966).
108. V. McKoy and J. P. McTague, private communication.
109. D. F. Heller and W. M. Gelbart, preprint entitled "Collision Induced Optical Properties of Simple Fluids: I. Short Range Electronic Distortion and the Density Dependent Dielectric Function."
110. See, for example, Ref. 3(*b*) and earlier papers cited therein.
111. (*a*) H. B. Levine and D. A. McQuarrie, *J. Chem. Phys.*, **49,** 4181 (1968); (*b*) *J. Chem. Phys.*, **44,** 3500 (1966); (*c*) Document No. 8807 deposited with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress.

112. J. Hirschfelder, C. F. Curtis, and R. B. Bird, *Molecular Theory of Gases and Liquids*, Wiley, New York, 1954.
113. B. R. A. Nijboer and L. van Hove, *Phys. Rev.*, **85**, 777 (1952).
114. B. J. Alder, H. L. Strauss, and J. J. Weis, private communication.
115. W. B. Daniels and R. K. Crawford, private communication to H. Strauss.
116. L. Jansen and P. Mazur, *Physica*, **21**, 193 (1955).
117. D. W. Oxtoby and W. M. Gelbart, preprint entitled "Depolarized Light Scattering Near the Gas-Liquid Critical Point."
118. See the well-known treatment of electric dipole fields in, for example, J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, 1966.
119. We are grateful to Prof. Robert A. Harris for suggesting this method of decoupling the multiple integral.
120. (*a*) M. Giglio and G. B. Benedek, *Phys. Rev. Lett.*, **23**, 1145 (1969); (*b*) A. B. Cornfeld and H. Y. Carr, *Phys. Rev. Lett.* **29**, 1, 28 (1972).
121. Obtained from σ_{HS} (Argon) = 3.16 Å by scaling with second virial coefficient data and solid-state lattice parameters. See Hirschfelder, Curtiss, and Bird, *Molecular Theory of Gases and Liquids*, Wiley, New York, 1954.
122. Calculated from the data of Ref. 9 and Equations (29) and (30) of Ref. 19.
123. From P. W. Schmidt and C. W. Tompson, in *Simple Dense Fluids*, H. L. Frisch and Z. W. Salsburg Eds., Academic Press, New York, 1968.
124. R. A. Pasmanter, R. Samson, and A. Ben-Reuven, *Chem. Phys. Lett.* **16**, 470 (1972).
125. R. A. Harris, *Chem. Phys. Lett.*, **19**, 49 (1973).

EQUILIBRIUM IN STELLAR SYSTEMS

R. H. MILLER

Department of Astronomy, Institute for Computer Research,
and
Comittee on Information Sciences,
The University of Chicago,
Chicago, Illinois.

CONTENTS

I. Introduction 108
A. Stellar Dynamical Preliminaries 109
B. Some of the Difficulties 111
C. Equilibria 113
D. Statistical Mechanics 114
II. Numerical Experiments on Equilibrium 115
A. The n-Body Calculation 117
1. The isothermal enclosure 118
2. Specularly reflecting enclosure 118
3. Division into subclusters 119
B. Runs with a Cold Box 119
C. The Box at Other Temperatures 120
D. Specular Reflector 121
E. Subclusters 122
F. Discussion of the Experimental Results 123
III. Pair Correlations 124
IV. The H-Theorem 126
A. The Inequalities 127
B. Consequences of Increasing H_B 128
C. Equilibrium of Stellar Systems 130
D. Rate of Increase of H_B 133
E. Summary 133
V. The Microcanonical Ensemble 134
A. Meaning for the n-Body Problem 137
B. "Thermodynamics" 139
C. Equilibrium of Star Clusters 140
D. The H-Theorem 141
VI. Concluding Remarks 142
References 143

I. INTRODUCTION

It is immediately apparent, on looking at a photograph of a galaxy, or of a cluster of galaxies, that these objects are not uniform. They are groups of stars or of galaxies that stand well separated from their neighbors. Equilibrium of a stellar dynamical system is *not* the uniform state of maximum entropy that characterizes equilibrium in most cases. Thus stellar systems provide a challenge: they represent what is quite likely the simplest kind of system occurring in nature for which the equilibrium state is not uniform; they are certainly the simplest one-component system known to display this property. The study of stellar dynamical systems promises to provide clues not only to the understanding of the beautiful objects we see in the sky, but also to things closer to home in which nonuniformity seems to be one of the principal attributes (biological systems?).

Stellar dynamical systems are discussed in this article. The model is an isolated system of n point particles, each of unchanging mass, interacting through Newtonian gravitation and obeying the Newtonian equations of motion (relativistic effects are omitted). This has the immediate consequence that the system admits a Hamiltonian description, and conserves energy, etc. We do not ask where the point particles came from, how they might have been formed, whether they actually last forever, what are reasonable initial conditions, and so on. Instead we inquire about the development of the n-body systems, and in particular, whether they possess nontrivial equilibrium solutions. The methods are primarily those of statistical mechanics: methods that promise an overview without requiring detailed solutions. But the usual price is paid—many of the obvious questions cannot be answered within this framework. A few special investigations amenable to this kind of treatment are presented in this paper; no attempt is made to provide a balanced overview of work in the field.

The inverse-square-law force of self-gravitating stellar systems is both long-range and strong at short range; collisions do not occur in the conventional sense. All particles are interacting all the time, with varying strengths of interaction. There is no screening. This feature makes it necessary to reexamine the fundamental assumptions that underlie the theory, to see which are still valid. For this reason, some of the work in rigorous statistical mechanics is of particular interest to stellar dynamics, because it causes the underlying assumptions to be stated explicitly and precisely.

With a restriction to finite numbers of particles, we recover the gravitational n-body problem. This problem has a long and distinguished history.

TABLE I

Properties of Some Typical Stellar Systems

Region	n (pc^{-3})	c_s (km/sec)	N	R (pc)	T_D (yr)	T_S (yr)	Coupling
Solar neighborhood	0.1	30			10^7	10^{13}	$2 \cdot 10^{-6}$
Galactic center	10000	200			$4 \cdot 10^4$	10^{11}	$2 \cdot 10^{-6}$
Elliptical galaxy	100	200	10^{11}	10^4	$4 \cdot 10^5$	10^{14}	$5 \cdot 10^{-7}$
Globular cluster			10^5	10			
Inner region	1000	10			10^5	10^8	$4 \cdot 10^{-4}$
Outer region	1	10			$4 \cdot 10^6$	10^{11}	$4 \cdot 10^{-5}$
Galactic cluster	10	0.5	10^3	1–2	10^6	10^6	0.04

The case, $n = 2$, yields the familiar Kepler problem, while $n = 3$ is the famous "three-body problem" that has stimulated so much work in both pure and applied mathematics. The approach of this paper is more along the lines of kinetic theory than of the n-body problem, however. Further, stellar dynamical systems are considered, rather than self-gravitating systems in general. This means that self-gravitating gaseous or plasma systems will not be considered, nor will systems in which both gas and stars play an essential part. Interaction with interstellar gas may be quite important; it can provide a dissipative mechanism that may turn out to be essential to understanding stellar systems. However, the difficulties of handling these more complicated systems, and the lack of reasonable observational data for comparison, indicate that a more fruitful approach will be to study n-body systems to see if they behave as do the systems we see in the sky.

A. Stellar Dynamical Preliminaries

This section contains descriptions of some typical stellar systems and of some of the concepts needed for later sections. Stellar systems exist in a wide range of sizes and total numbers of stars from single stars to binary or other small (dynamically bound) groupings, to associations (up to 100 stars), to galactic and globular clusters, to dwarf galaxies of 10^8 to 10^9 stars, to normal galaxies of 10^{11} to 10^{12} stars. Some "supergiant" galaxies may have 10^{14} to 10^{15} stars, but this seems to be an upper limit to the size of known stellar systems. Larger groupings seem to fragment into objects of smaller size, usually galaxies. It is not known how far these hierarchies extend. Typical properties of some of these systems are listed in Table I. There the number density (number of stars per cubic parsec, 1 parsec = 3×10^{18} cm, or about 3.26 light years), the rms velocity of random motions (c_s, in kilometers per second), the total number of stars, N, typical size

of the systems (in parsecs), two different time scales, and a coupling parameter, which will be explained later, are given.

An elementary consequence of the n-body equations of motion is the scalar form of the "virial theorem": a transformation of the equations leads to the Lagrange-Jacobi identities:

$$\frac{1}{2}\frac{d^2I}{dt^2} = 2K + \mathscr{V} \tag{1}$$

where I is the trace of the inertia tensor for the particle configuration, K is the total kinetic energy, and $\mathscr{V}$ is the total potential energy (reckoned from $\mathscr{V} = 0$ for infinite dispersal of all particles). The coefficient, 2, on K is appropriate to the inverse-square-law force. The "virial theorem" results from the assertion that suitably long time averages give a zero result for the left-hand side of (1), so the time average of the right-hand side is zero.

A second property of self-gravitating systems is that a dynamical time scale is defined by $T_D = (4\pi G\rho)^{-1/2}$; this may be seen on dimensional grounds, or by solving for the free-fall collapse of a uniform self-gravitating sphere. This is the entry labeled T_D in Table 1. The dynamical time scale gives the order of magnitude of the time in which properties develop as a direct consequence of the equations of motion. In the solar neighborhood, T_D is nearly the time required for the phase of an oscillation perpendicular to the galactic plane to advance through one radian. In rotating systems, the angular velocity is typically around T_D^{-1}, and with a nonrotating cluster, a typical star can cross the cluster dimensions in a "crossing time" that is about the same as T_D.

A third property of typical systems is that any two particles seem weakly coupled. A measure of the strength of the coupling is the ratio of the potential energy of near neighbors divided by the kinetic energy of their relative motion,

$$\text{"coupling"} = \frac{Gm}{d_n c_s^2} \tag{2}$$

where d_n is the mean nearest-neighbor separation. Through the virial theorem, this coupling can be seen to be related to the ratio of the potential energy of the gravitational interaction between a star and its nearest neighbor to the total potential energy with which it is bound to the cluster. This parameter is listed in Table I; for most cases, it is very small. Thus the system is in some sense "weakly coupled." The interaction between any pair of stars is quite a small fraction of the total interaction with the entire system. This leads to a situation in which, for most purposes, a star may

be regarded as moving in a "smooth potential," or the mean field, and in which binary encounters are a very small perturbation to the mean-field motion.

A fourth property is the usual relaxation time, computed from summing deflections from the mean-field orbit that arise from binary encounters. Because of the inverse-square-law force, this calculation is just like that of multiple coulomb scattering for charged particles in matter. Because the interaction is binary, the relaxation time is typically of the order of T_D/coupling, although the precise relationship is somewhat more complicated. We have no need for the analytic form,[1] but the order of magnitude is often very long—10^{11} to 10^{13} yr. Systems like this have not had a chance to relax appreciably during the age of the universe, which is estimated at about 10^{10} yr. This relaxation time scale is sometimes referred to as a secular time scale, and is listed in Table I as T_S. It provides a measure of the time during which a mean-field orbit for a star is presumed to be a good approximation.

A fifth property of stellar systems is that the infinite uniform medium is gravitationally unstable. This instability is like an imaginary plasma frequency, and is distinct from the notion of a mean mass density that will "close the universe." In the absence of rotation, this so-called "Jeans instability" occurs in uniform media of mean-square velocity c_s^2 for all wavelengths that satisfy the inequality[2,3]

$$\lambda_J^2 > c_s^2 T_D^2. \tag{3}$$

Most stellar systems have system dimensions near the Jeans unstable wavelength. Because there is no screening of the gravitational interaction, as there is of the electrostatic interaction in a plasma, there is no analogue of a Debye length; the system dimensions must be inserted wherever a Debye length would usually be invoked as an upper limit cutoff. This has been done, for example, in recent reexaminations of the conventional relaxation problem[4]; it provides a natural long-range cutoff to replace the older rules, which had an ad hoc flavor.

B. Some of the Difficulties

The weakness of the coupling suggests that the usual techniques of statistical mechanics should be readily applicable to stellar systems. Unfortunately, the situation is nowhere nearly so nice. Even the apparently simple matter of collisions is complicated; because there is no screening, and because of the long range of the interaction, every particle of the system is undergoing a "collision" with every other particle all the time. The usual notion that collisions are of short duration or that the time

during which particles collide can somehow be separated from the rest of the time does not carry over into stellar dynamical problems. Collision operators have complicated long-time properties, and many collisions overlap in time.

Some years ago, Chandrasekhar studied stellar dynamics from the viewpoint of stochastic processes,[5] as if the motions were similar to Brownian motion. This remarkable series of papers terminated with the observation that the autocorrelation of the force at a point in the system dies out very slowly in time, as $1/t$, rather than as the exp $(-t)$ that would be expected for a Markov process. This result has been confirmed by later workers, who have approached the problem from different points of view.[6,7] The long-time memory of these systems essentially vitiates this approach; attempts to integrate over the force autocorrelation lead to divergent integrals. The long-term memory has been interpreted as indicating a nonergodic system.[7]

The origin of this difficulty is not clear. It may be related to the use of an infinite system. However, finite systems should also show the long-term memory. Presumably, it cannot be traced to anything as simple as the Jeans instability of the infinite system. Nonetheless, the Jeans instability, and the observed division of stellar systems into fairly well-defined and well-separated clusterings indicates that the problem must be approached as a problem in finite, nonuniform systems if equilibrium is to be studied.

These properties (long-range force, no screening) have several immediate consequences. First, there can be no thermodynamic limit in the strict sense; attempts to pass to a limit of infinite volume and infinite number of particles (even if N/V is held constant) must run into trouble with the Jeans instability. It might still be hoped that the number of particles can be taken large enough that the "graininess" will be suppressed $O(1/n)$. However, there are so far no proofs that this will be possible.

A further consequence is that the usual thermodynamic distinction of intensive and extensive parameters is not possible. Energy is not extensive, for example. A cluster cannot be cut into halves, which are then separated. This would lead to a qualitatively different system, not simply to a smaller sample of the original cluster. Stellar systems cannot be subdivided: a small part of the system cannot be considered as an almost isolated system that exchanges bulk properties across the boundaries that separate it from the rest of the system, as is frequently imagined in thermodynamical or statistical mechanical arguments. The entire system is one large, coherent whole. Even if a diathermic wall is considered, the gravitational interaction will penetrate the wall. This means that the canonical ensemble and grand canonical ensemble are not useful. The only one of the usual ensembles of

statistical mechanics that can be used is the microcanonical ensemble. The system must be considered as a whole; particle number, total energy, centroid momentum, and angular momentum are all conserved. We have thus rejected, at the outset, many of the tricks that have rendered some problems tractable in statistical mechanics.

The numbers of particles that exist in stellar systems are rather few for some of the usual arguments about differentials of coordinate or velocity space in statistical mechanics. Even in the six-dimensional μ-space, division of the coordinate axes into 100 divisions along each axis would leave only one star per phase cell, a rather crude division. Some of the limiting processes usually invoked in statistical mechanics must be used with caution.

C. Equilibria

It is not at all clear that there are equilibria for stellar dynamical systems. But there must be metastable equilibria, in the sense that states should exist that, although they are not strictly equilibrium states, still change so slowly that they represent equilibria for all practical purposes. For example, it is known that stars escape from clusters,[1] but clusters are observed in nature, so a star cluster represents a practical, if not a mathematically exact, equilibrium configuration for a group of stars. Questions of stability follow discovery of equilibria, and are not considered here.

In Section IV, the consequences of trying to define equilibria, or evolutionary trends, by means of the H-theorem are investigated, with the conclusion that the H-theorem cannot be used to find equilibria for stellar dynamical systems. It is not surprising that this should occur; the notion that evolution (under the second law) leads to "states of maximum disorder" seems to be contradicted for stellar systems. In fact, the requirement that the systems obey the Liouville theorem is shown to indicate that evolution is toward states characterized by a certain kind of order.

The time scales for evolution found in Section IV are secular time scales; they are so long that the entire discussion takes on a vaguely unreal feeling. Obedience to the n-body equations of motion is demanded for all time; the first integrals of motion and the Liouville theorem must hold forever. But evolution, as the system seeks an equilibrium, proceeds so slowly that the entire age of the universe—even the expected lifetimes of the longest-lived stars—is insufficient for the evolution to have developed to any significant degree. Concern for some of these points—indeed, questions of ergodicity and reversibility—take on a flavor of mathematical fine points. It may be that some of the features that have been ignored, such as interaction with interstellar gas clouds, may provide sensible resolutions to these problems. Until such questions are straightened out,

acceptance of these points at face value seems to be the only way to proceed.

A similar set of questions arises with regard to treatment of systems as isolated. It is impossible to isolate stellar dynamical systems from each other completely. An extreme view would even deny the possibility of constructing ensembles because there is no place in the universe where the other ensemble members might be located to assure independence. Although it is true that different systems interact through the long-range forces, the principal effect must be a global effect between two clusters; details of the internal structure of each are communicated by higher-order multipole terms, which die off so fast with increasing distance that they should give no trouble. Interaction between clusters should not seriously affect the internal dynamics of either.

Thus the conclusion, from both Sections IV and V, that there is no equilibrium state, is not a serious matter. The realistic metastable equilibria, which are states that dissolve slowly, escape detection by the methods used. The conclusion stands on the basis of unrealistically long time scales; for the times over which these systems could have existed, a mean-field (gas) description is probably as useful as any technique now available. But even that description must allow for the fact that stellar systems have a highly correlated structure. Some aspects of that problem are developed in Section III.

D. Statistical Mechanics

Quite apart from questions of the existence of equilibria, there are questions whether statistical mechanics is applicable. Several viewpoints are possible, ranging from the denial by Kurth[8] that the methods of statistical mechanics can be used at all to the view that statistical mechanics offers some useful techniques for treating what is basically a very complex problem in mechanics. The most direct way to find out whether statistical mechanics is applicable comes through attempts at working out some examples. Comparison with observation, or with other methods of studying realistic systems, indicates whether it is applicable. Attempts to use the techniques should not lead to contradictions or to ridiculous results. The investigations reported here stand up under these tests, but they also indicate that considerable care is required in order to avoid some obvious pitfalls.

There are other difficulties beyond those discussed so far; for example, the entire matter of observations of the systems under study. With stellar systems, observations provide only a "snapshot" of the instantaneous microstate of stellar system, and an incomplete snapshot at that. They

never provide the averaged macroscopic properties that are usually treated in statistical mechanics. The only truly macroscopic property we see is the hierarchical structure of clusterings. This situation is the opposite of that usually prevailing in statistical mechanics, where even the shortest observation typically requires many relaxation times to accomplish, and where interactions with walls, etc., provide ways of admitting violations of the equations of motion that are slow, but still fast compared with the time required to observe the system.

Notwithstanding these evident difficulties, many workers have been led by the apparent power and generality of statistical mechanics and of thermodynamics to try to apply these methods to problems in stellar dynamics and in related areas.[9–13] One feature of self-gravitating systems that is particularly attractive is the "negative specific heat," which arises in systems so large that gravitation provides a significant fraction of their internal energy. This has led to predictions of a "gravothermal catastrophe."[10,12] But one can easily be led to formulations that violate the equations of motion, or to other difficulties.

The parts of statistical mechanics that seem to survive all these special requirements are the microcanonical ensemble, along with the notions that the a priori probability of certain sets of states is proportional to the volume accessible to those states in the $6n$-dimensional phase space, Γ (the microcanonical ensemble has infinite volume accessible), and that additional information restricts the volume available in Γ.

II. NUMERICAL EXPERIMENTS ON EQUILIBRIUM

One of the difficulties associated with astonomy is that it is impossible to do experiments on the systems of interest; they can be observed only in a rather passive way. It would be very nice to go around to look at a system from a different direction, or to kick it to see if it would bounce. We either have a frustrating view of the details of a system (our own Galaxy, for example), such that large-scale structure can only be inferred by very indirect arguments, or we can see only the grand structure and must guess at the details (as with other galaxies). We can see the entire forest or the trees from which it is made, but never can attain a balanced view of both. (Some astronomical problems also involve hostile environments that cannot be duplicated in the laboratory.) The dynamical development of star clusters or galaxies occurs on a time scale of millions of years so that our observations, imperfect as they are, at best represent a snapshot of instantaneous particle positions and velocities. We cannot observe the systems long enough to see how they will develop or evolve.

Essentially the only recourse available, to serve as a guide and check on theory in the sense that experiments fill this need in other physical sciences, is through computer simulations, or numerical experiments. The experiments reported in this section represent a tentative step toward seeing how "laboratory" stellar systems would respond to being subjected to "thermodynamic experiments." They were originally designed with the hope of constructing a kind of Carnot engine, in which a stellar system would be the thermodynamic fluid. However, the results were so far removed from "thermodynamic" behavior that it was out of the question to interpret them in thermodynamic terms. For example, it was not possible to formulate an operational definition of a thermodynamic temperature by the usual method of allowing the system to come to equilibrium with a heat reservoir. But that is already a useful guide; the attempt to understand these results has led to some insights into the properties of stellar systems. The effects discussed in the remainder of this paper were first found experimentally; the reasons why the experimental systems behaved as they did emerged only after a fairly careful study of the premises underlying some attempts to construct a thermodynamics of self-gravitating systems. The experimental techniques and results are described in Sections II.A through II.E; the interpretation in terms of permitted violations of the n-body equations of motion is presented (somewhat out of the logical order for the paper) in Section II.F.

Numerical experiments share with all other experimental techniques the need for the experimenter to demonstrate that his results actually imply the conclusion he has inferred; that the results are not simply dominated by some form of experimental error, and that the measurement made is relevant to the problem under study. (The same situation obtains with theoretical developments as well, but there it usually receives less attention.) Several of the points that arise in this connection will be mentioned without going into details to give some of the flavor of the problem.

The experimental tool used is an n-body calculation, with inverse-square forces between each pair of particles. The Newtonian equations of motion are integrated for this n-body system as accurately as computational techniques allow. The need to compute so many force terms limits the number of particles that may be used in computer simulations to a few hundred; the experiments reported here are based on still smaller numbers of particles. Systems are started from arbitrary initial conditions; it is hoped that the observed properties do not all derive from a bizarre initial condition. The details of the n-body calculation are complicated by the need to handle close encounters (hyperbolic Keplerian orbits), formation of moderately long-lived binaries (bound Keplerian orbits), and a long-range force such that the interaction of every pair of particles

must be taken into account, no matter how large their separation. All these features lead to a system that is remarkably unstable from a numerical point of view[14]; indeed, the gravitational n-body calculation is probably the most unstable of any of the computations that are seriously used in attempts to understand physical systems. The numerical instability arises because neighboring phase trajectories separate exponentially in time. This phenomenon is typical of mixing systems; the gravitational n-body problem seems to be a mixing system. For the experiments reported here, the numerical difficulties were avoided by using a variety of initial conditions, restricting each calculation to a time short enough to keep the numerical error growth under control.

Few particle experiments cannot approximate thermodynamical or statistical systems at all well; there is no clear-cut recipe to construct a distribution function from the experimentally available parameters, notwithstanding the complete information on the state of the n-body system that exists within the computer. One representative member is simply not an ensemble.

It is difficult to demonstrate a thermodynamic limit for any computed n-body system, but Alder[15] has shown that it is possible to construct a computed system that displays a thermodynamic limit (a hard-sphere system), even though this occurs only at high- or low-density limits with several hundred particles. Such large numbers of particles are not accessible for the kind of gravitational n-body calculation described here; furthermore, it is not clear what the terms "high- and low-density limits" mean in stellar dynamics.

There are also conceptual problems, such as building massless enclosures, whose only function is to contain the particles. Even the notion of enclosures that would contain a system of stars is rather bizarre. The abstraction of an isolated n-body system may be too unphysical; we have little appreciation of the importance of neighbors on the internal dynamics of a cluster. In nature, there is a complex interplay of many effects, but that is precisely the reason for running the experiments.

A. The n-Body Calculation

The routines used for the calculation were conventional n-body routines, like those described some years ago by von Hoerner.[16] Most of the special methods now commonly used to speed up the calculation[17] were not incorporated; fast-running calculations were not as important for this project as was a routine that could readily be adapted for special purposes.

Several different calculations were started from a variety of initial conditions, each run for a fairly short time—at most, a few crossing times of the initial condition. A kind of ensemble average was required, as the

system was not expected to be time independent. The starting conditions were simple configurations with the kinetic and potential energies satisfying the virial theorem.

Most of the calculations were run using 32-body systems. Units were chosen so the gravitational constant $G = 1$, and the particle mass $m = 1$ (for each particle; all particles have the same mass). A variety of particle masses complicates the interpretation. Thirty-two particles is certainly very few for "thermodynamic" experiments, but since it might be hoped that finite-system effects should decrease relative to thermodynamic effects by terms $O(1/n)$, qualitative thermodynamic effects should emerge even with this small a number of particles.

1. *The Isothermal Enclosure*

For experiments mimicking an isothermal enclosure, the system was placed in a box, which works as follows. After each integration step, each particle is checked to see if it is outside the box. Any particle outside the box is reprojected into the box; its new kinetic energy is random, exponentially disturbed with mean value T. Changes in kinetic energy (and the change in angular momentum) are tallied. The particle is reprojected from the location it had when it was discovered to be outside the box. A cubic box, with boundaries at ± 1 in all three directions, was used.

The box is endowed with a "'temperature," T, that governs the mean kinetic energy of the reprojected particles. It is also endowed with a "heat capacity," according to which its temperature can change in proportion to the net energy interchange between the stellar system and the box. Most experiments were run with a rather large heat capacity, and the resultant temperature changes (for the box) were negligible. This approximates the isothermal enclosure.

It would be possible to design an adiabatic enclosure for these systems; however, adiabatic enclosures were not used in the experiments reported here.

2. *Specularly Reflecting Enclosure*

Several experiments were conducted with a specularly reflecting spherical enclosure. These produced results that were quite different from those with the isothermal enclosure; the results are described in Section II.D.

As with the isothermal enclosure, each particle was tested after each integration step to determine whether it were outside the (unit) sphere. Any particle that was outside had its radial component of velocity reversed, while retaining the plane of the motion. Its position was unaltered. This change separately conserves kinetic energy, potential energy, and angular

momentum about the center of the sphere for each collision; no net energy transfer to the enclosure can take place.

3. *Division into Subclusters*

All the methods for placing stellar systems into enclosures or to test their response to "probes" seem unnatural. A more natural approach is to monitor the transfer of energy (or of other properties) among various constituents of the stellar system. This can be studied easily in n-body calculations by grouping the particles into subsets according to some simple label. Experiments conducted by this strategy have the advantage of not requiring any special boundary conditions or other unusual features; the experiment consists of making appropriate tallies of the information available at any time throughout an otherwise ordinary calculation.

The details of the subgroupings used and the quantities tallied are described in Section II.E, along with the results of the experiments.

B. Runs with a Cold Box

A particularly simple condition results if the box temperature is zero. Any particle that ventures outside the boundary is stopped, its kinetic energy is removed, and the particle is then released to fall from the point at which it was caught. This practically assures that the particle will fall back through the most dense part of the cluster, where it can interact very strongly with the remaining cluster members. This "cold box" condition produced some very interesting results.

The energy delivered to the box is shown as a function of the time for "cold box" runs in Fig. 1. The kinetic energy of the initial state was about 250, the potential energy about -500, and the total energy about -250. The time units of the abscissa were arbitrary, but one unit was approximately the time required for a particle with the mean kinetic energy of the initial condition to cross the box. It is customary to quote times in terms of a "crossing time," the time required for a particle with average properties to cross the cluster. As a cluster develops, the crossing time may change. The time unit of the abscissa is essentially one crossing time of the initial cluster.

The remarkable features of Fig. 1 are the rate at which energy is given up to the box and the fact that the rate does not diminish appreciably as the process continues, even though the total amount of energy transferred is quite large (four times the total kinetic energy in the initial state). The solid curve represents one calculation; the dashed curve represents another calculation from similar starting conditions for comparison purposes. The two curves give some idea of the reproducibility of the results from experiment to experiment.

Fig. 1. Energy transfer from a 32-body system to a cold enclosure. The total initial kinetic energy was about 250. The time units on the abscissa were nearly a crossing time of the initial state. The two tracks represent two distinct calculations.

During these runs, the virial theorem remained approximately satisfied, without any allowance for the effect of the boundaries. The ratio of total potential to total kinetic energy of the particles was near the expected -2 most of the time, but showed fluctuations toward -1 whenever an unusually close encounter occurred. The star cluster got smaller and smaller as more and more energy was given up to the walls. This means that the duration of a crossing time decreased, so around $t = 6$, one unit of t represents about 10 crossing times. As seen from the cluster, the rate of giving up energy to the enclosure decreases rather rapidly. The cluster at the end of the experiment is quite different from the initial cluster. At the end, it is quite compact, presumably near the center of the box, with very large kinetic energy. Its properties may be reasonably inferred from the virial theorem.

C. The Box at Other Temperatures

A second series of experiments was run in which the temperature of the enclosure was not zero. Since the temperature of the box gives the mean kinetic energy with which stars are reprojected into the interior after striking the wall, it is convenient to quote the temperature in units of the mean kinetic energy of the particles in the initial condition (about 8, typically, in the $G = m = 1$ units of the calculation). With the enclosure

temperature at or below the mean kinetic energy of the particles (in the initial state), the stellar system gave up energy to the enclosure at about the same rate as it did to the cold box. The energy transfer started immediately. The behavior was similar to that observed with the cold box.

Only when the temperature became twice as large as the mean particle kinetic energy did a sensible change in behavior appear. Then the stellar system exchanged energy with the enclosure, undergoing rather large fluctuations in the energy balance. After a while, the system changed over to behave as if it were in a cooler enclosure, giving up energy to the enclosure.

The stellar system tends toward a state in which it looks as if it and the enclosure have the same temperature, but it then starts to transfer energy to the enclosure, as in the earlier cases for which the box temperature is the same as the mean kinetic energy per particle. When the transfer finally begins, the rate of energy transfer to the enclosure is perhaps half that with the cold box. It is not clear what determines the rate of energy transfer once the transfer begins.

Presumably, if the enclosure were very much "hotter" than the initial stellar system, the readjustment of the potential energy might not occur, so the system would not start the transfer. Experiments have not been carried out into this range, which might require box temperatures 10 times the mean initial kinetic energy per particle.

There is a question as to the effect of the fluctuations in precipitating the onset of the energy transfer. Certainly the fluctuations must play some part in triggering the transfer, but, even in a special set of experiments designed for the purpose, it was impossible to identify cooperative effects in the fluctuations. This makes it difficult to attribute the onset of the transfer to the fluctuations, especially since the cluster tends toward a state which is ripe for the transfer before the transfer actually starts. The fluctuations may play a part in increasing the (magnitude of the) potential energy. Additional experiments would be required to answer this question.

D. Specular Reflector

The experiments carried out with a specularly reflecting sphere behaved in a qualitatively different way. No example tried thus far has shown a behavior like that associated with the energy transfers of the cases run in an isothermal enclosure. The specular boundary condition prohibits energy transfer to the enclosure, but the transfers in the isothermal enclosure were associated with a shrinking of the cluster and an increase in the kinetic energy that would be easy to identify if they did occur in the specular enclosure. Energy transfers can still take place among various constituent parts of the cluster; we shall continue to use the term "transfer" to

describe the general behavior of shrinking and "heating" of the cluster, even though it no longer means energy transfer to the enclosure. Of course, the specular reflector conserves energy and angular momentum at each collision and thus is quite different from the boundary conditions represented by the boxes described in the previous sections. Nothing surprising happened in the runs with the specular reflector, except for the sharp contrast shown with the earlier runs with the box. The virial theorem remained reasonably satisfied. The surface-pressure term was appreciable (up to 20%).

Experiments were tried with a variety of initial conditions. All had zero total angular momentum, with the mass centroid and velocity centroid on the origin. Initial particle positions and velocities were chosen by using pseudorandom number generators, with the velocity and density profiles approximating certain standard solutions to the self-consistency problem for self-gravitating systems. Ratios of central to peripheral densities of 1, 30, 709, and 5000 were tried. Sampling density distributions differing by factors of 700 or 5000 are not very meaningful with only 32 particles—there are not enough particles to probe the density variation. With these large density ratios, the particle system does not know it is in an enclosure. A few particles did strike the sphere in these experiments—but they were escaping particles that were set back into the cluster by the reflection.

E. Subclusters

The mechanism by which "negative specific heat" is argued to lead to the formation of core-halo structures in the "gravothermal catastrophe"[10] is based on energy transfers between "hotter" and "cooler" subsets of particles within a given cluster. The terms "hotter" and "cooler" refer to particles with high and low kinetic energy, respectively. One way to mimic these notions in a machine calculation is to separate the particles of an ordinary calculation into two classes according to their kinetic energy. As indicated earlier, this set of experiments makes use of a normal calculation for a star cluster, and is free of the unnatural boundary conditions of the other experiments. Whenever a fairly complete summary of cluster properties is run, the particles are sorted into two subsets according to kinetic energy. Along with the usual cluster properties, the kinetic and potential energy of each subcluster is tallied as if the rest of the particles did not exist; the potential energy of interaction of the two subclusters is also tallied. A given particle may move freely from one to the other of the two subsets. The separation applies only to the tallies; no distinction is made for the integration.

This model appears to be very much in the spirit of discussions on "negative specific heat" and of the "gravothermal catastrophe."

No strong tendency to transfer energy from the "hot" to the "cool" subcluster is evident in these experiments. The low-kinetic-energy subcluster nearly obeys the virial theorem as if the other subcluster were not present.

An alternative way of looking at these experiments that is free of some of the arbitrariness of division into two subclusters is to study the distribution of particle kinetic energy at various times. This distribution definitely did not show a tendency to segregate into extreme examples of high and low kinetic energy; high kinetic energies increased sharply during close encounters, but apart from this effect no trend toward relatively distinct subsets of high- and low-kinetic energy particles appeared.

These experiments are consistent with the maintenance of an isothermal structure. Although the number of particles included in the experiment was not very large, it was not possible to detect a tendency to split up into two (or more) populations with different "temperatures."

F. Discussion of the Experimental Results

Among the various computer experiments, that with the isothermal box is clearly the anomalous case. Both the examples with the specular reflector and those in which the system was regarded as if made up of two different subsystems showed no tendency to shrink into a small, "hot" cluster or to subdivide into a small, "hot" system that coexists with another extended, "cool" system ("core-halo" structure).

The essential difference between the examples run with the isothermal enclosure and the other cases seems to be that violations of the equations of motion can occur at each encounter with the isothermal enclosure that permit changes in the allowable phase volume in Γ. By contrast, neither the specular reflector nor the subdivision of the particles into classes permits this. The subdivision manifestly leaves a Hamiltonian system. The specular reflector can be described as a Hamiltonian system by introducing an infinite (positive) potential at the position of the wall. Thus both of these systems have a Liouville theorem in Γ. For them, the condensation into a compact system does not occur. There is no such constraint on the problem with the isothermal enclosure.

A collapse, if it is to occur, must do so on a secular (relaxation) time scale, and not on a dynamical time scale (Section IV). Unfortunately, n-body calculations are not well suited to an examination of this question since the secular and dynamical time scales are not well separated for systems that contain fewer than about 1000 particles. However, this feature is also somewhat advantageous: processes that occur only on a relaxation time scale can be studied in n-body calculations because that time scale is short enough to be within the reach of practical computations.

The time scale of energy transfer to the enclosure in the experiments in an isothermal enclosure may have been dominated either by dynamical or by relaxation time scales, or by still another (undefined) time scale. However, it is suggestive that the rate of energy transfer did not change with the dynamical time scale of the cluster as the cluster shrank. As the cluster got smaller in the configuration space, and the dynamical time scale got shorter in the units of the calculation, the rate of energy transfer stayed about the same. It would have decreased if measured in terms of the dynamical time scale. It would not decrease nearly as strongly if measured in terms of a relaxation time scale. The energy transfer, and the shrinkage of the residual cluster, seem to have been dominated by the relaxation time scale, as expected.

It was not possible, in these experiments, to define a thermodynamic temperature of a stellar system by the operational method of placing the system in contact with a foreign body, and allowing the two to reach an equilibrium. No equilibrium was reached. Again, this can be attributed to the violations of the equations of motion and of the Liouville theorem that are permitted. The stellar system, if allowed to interact with a "heat bath," is no longer a Hamiltonian system by itself. It need not obey the Liouville theorem in Γ, and can undergo a collapse like the "gravothermal catastrophe." It might be possible to construct the "heat bath" in such a way that the total system is describable by a Hamiltonian; there must then be some kind of Liouville theorem for the total system. But the phase volume accessible for the stellar component, viewed as a separate subunit of the total system, need not be conserved. It thus appears to be impossible to design a system that would permit an operational definition of "temperature" for a stellar dynamical system.

III. PAIR CORRELATIONS

Stellar dynamical systems show a strong pair correlation, which appears as a tendency of two particles to avoid each other in the six-dimensional μ-space.[18] It results from a long-range interaction that is strong at short range, and is not peculiar to attractive inverse-square-law forces. If two particles were near each other in μ-space, they would be close together in configuration space, so they would have a large force acting between them. But they would have very little relative velocity. The large force would generate a large relative velocity in a very short time, so the two particles could not remain near each other in the μ-space very long.

A pair correlation should be described in the 12-dimensional phase space of particle pairs; the correlation found in stellar systems can be described as an exclusion of more than one particle from a region in the

six-dimensional μ-space. The correlation extends to all orders of reduced distribution functions (in the BBGKY hierarchy), causing all reduced distribution functions above the single-particle function to vanish whenever two (or more) particles are near each other in the μ-space. This

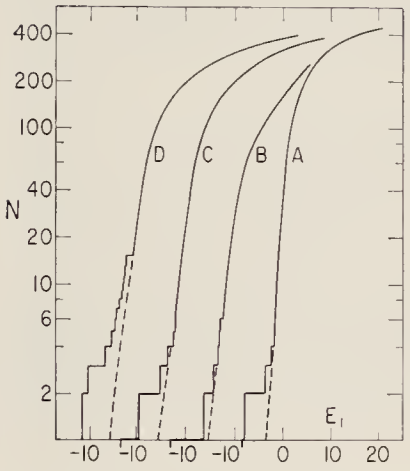


Fig. 2. Experimental results on pair correlations. Histograms of log N vs. E_1, where N is the cumulative number of pairs of particles for which $E_{pr} < E_1$. The individual plots are displaced by 10 units along the abscissa for clarity. The dashed lines are straight-line extrapolations to $N = 1$ of the linear portion of each histogram. Plot B is the same experiment as Plot C, but at a slightly earlier time (0.3 crossing times instead of 0.4). Plot D was obtained from the data of Plot C by permuting the r_{ij} associated with a given v_{ij}, thereby destroying the pair correlation. The value of E_{pr} for $N = 1$ is -126 for Plot D. Note that this plot is quite similar to Plot C above $N = 15$, but tends toward negative E_{pr} for $N < 15$. The most spectacular difference is at $N = 1$, of course. Plot A is a different experiment (different initial conditions) for comparison purposes, run to 1.5 crossing times.

indicates a very strong correlation, with all the reduced distribution functions (and the n-particle function as well) vanishing on appropriate hypersurfaces through their respective phase spaces. In particular, the pair correlation does *not* die off with increasing particle number; it is of order unity irrespective of the number of particles in the system, and it appears in the description of the reduced distribution functions at the same order as the reduced distribution functions themselves. This property is likely to limit the usefulness of the BBGKY hierarchy in stellar dynamical problems.

The pair correlation was found by means of numerical experiments[18,19] (a different set of experiments from those described in Section II). On the basis of experimental evidence, the variable E_{pr} was singled out to describe the pair correlation; E_{pr} is the energy the pair of particles would have as a Keplerian pair if there were no other particles in the system:

$$E_{pr} = \tfrac{1}{2}v_{12}^{\,2} - \frac{2G}{r_{12}} \tag{4}$$

The quantity E_{pr} is not conserved. The logarithm of the cumulative number of pairs with $E_{pr} < E_1$ is plotted as a function of E_1 in Fig. 2;

results of four different experiments are displayed. The nearly straight portions below $N = 60$ lend weight to the notion that E_{pr} may be the appropriate variable with which to measure the correlation, and suggest that the probability of reaching a negative value of E_{pr} is proportional to $\exp(-E_{pr}/E_0)$ for $E_{pr} < E_2$. The "saturation" for $N > 100$ results from the shape of the cluster; it is independent of the pair correlation. Values of E_0 obtained from Fig. 2 are in the range -1 to -2, and values of E_2 obtained by extrapolating the straight lines to $N = 496$ (the number of pairs in a 32-body system) are in the range $+4$ to $+7$. The total energy in the cluster, in similar units, was about -220. Because E_0 is negative, this is the reverse of the usual exponential distribution; more negative values of E_{pr} are less probable.

The pair correlation effect can be dramatically displayed by destroying the pair correlation without otherwise affecting the cluster; this is most easily done by computing E_{pr} with the kinetic energy of one pair and the potential energy of some other pair. Curve D of Fig. 2 was obtained by permuting the r_{12}'s associated with the given v_{12}; the departure from a linear relationship below $N = 15$ is clear in this plot. The pair correlation is limited to the (spatially) closest pairs. The effect is by far the strongest with the closest pair, which has $E_{pr} = -126$ in Curve D. Curves C and D represent identical clusters, one with the correlation (Curve C) and one without.

It is interesting to speculate[19] that the pair correlation in stellar dynamical systems, because it behaves like an exclusion, might be used to define a cell size for the phase space, leading to equilibrium distributions similar to Fermi-Dirac distributions for nonzero temperature. This is possible because the volume of μ-space accessible to a Keplerian pair with $E_{pr} < 0$ is finite. With reasonable guesses for the appropriate value of energy to assign to the Keplerian pair, it turns out that typical stellar systems are essentially nondegenerate. The reader should be warned that this division of the phase space into cells is pure conjecture, suggested by the nature of the observed pair correlation, and has no theoretical basis whatever at the present time.

IV. THE *H*-THEOREM

It is unlikely that there is an *H*-theorem for stellar dynamical systems. Prigogine and Severne[7] have shown that there is not an *H*-theorem for infinite uniform self-gravitating systems, but such systems are known to be unstable by arguments other than the *H*-theorem (the Jeans instability). The situation is less clear with finite systems, which we are forced to consider in stellar dynamics. This section is devoted to an exploration of

some of the consequences of an H-theorem, supposing it to hold. The principal tool is the set of inequalities commonly used in information theory.

A. The Inequalities

To facilitate direct application of theorems from information theory, the distributions are probability densities, all normalized to unity: the n-particle distribution function is defined in Γ, it is taken to be symmetric in all particle indices, all particle masses are taken to be equal, and the reduced distributions are simply marginal projections of f_N:

$$\begin{aligned} f_1(1) &= \int d^{3(N-1)}x\, d^{3(N-1)}v f_N(1, 2, \ldots, n) \\ f_s(1, 2, \ldots, s) &= \int d^{3(N-s)}x\, d^{3(N-s)}v f_N(1, 2, \ldots, s, s+1, \ldots, n) \end{aligned} \tag{5}$$

with the usual convention that the integrals are over the variables that appear in f_N but do not appear in f_s. Note that these definitions differ in the normalization from those usually used in statistical mechanics. The functions

$$\begin{aligned} H_B &= -\int d^3x\, d^3v f_1 \log f_1 \\ H_G &= -\int d^{3N}x\, d^{3N}v f_N \log f_N \end{aligned} \tag{6}$$

are then the same as the definitions of "entropy" used in discussions of information theory.[20] The subscripts B and G signify that these are equivalent to the Boltzmann and Gibbs H-functions, respectively.

The function, H_G, is a constant of the motion as a consequence of the Liouville theorem, itself a consequence of the n-body equations of motion. Because stellar dynamical systems are assumed to be isolated, and to have no interaction with walls, etc., obedience to the n-body equations of motion is the principal postulate of the entire discussion.

The basic inequality is

$$nH_B \geqslant H_G \tag{7}$$

with equality if and only if the n-particle distribution can be written as a product of single-particle distributions:

$$f_N(1, 2, \ldots, n) = f_1(1) f_1(2) \cdots f_1(n) \tag{8}$$

Proofs for all inequalities to be used here are straightforward.[20] The special set of conditional distributions is also useful:

$$\begin{aligned} c_2(2 \mid 1) &= \frac{f_2(1, 2)}{f_1(1)} \\ c_3(3 \mid 1, 2) &= \frac{f_3(1, 2, 3)}{f_2(1, 2)} \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \tag{9}$$

These describe the conditional probability to find particle 2 at a phase-point $(x(2), v(2))$ when it is known that particle 1 is at $(x(1), v(1))$, and so on. The set of "conditional entropies"

$$\begin{aligned} H_2(2) &= -\int d^6x(1, 2)\, d^6v(1, 2) f_2(1, 2) \log c_2(2 \mid 1) \\ H_3(3) &= -\int d^9x(1, 2, 3)\, d^9v(1, 2, 3) f_3(1, 2, 3) \log c_3(3 \mid 1, 2) \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \tag{10}$$

can then be defined. Notice that the conditional H of particle 2, $H_2(2)$, is the average of $H_B(2)$ for each value of the coordinates of the first particle (1), weighted by the probability of getting that particular value for the coordinates of particle 1. Similar interpretations apply for the higher-order conditional H's. In terms of these quantities,

$$H_G = H_B(1) + H_2(2) + H_3(3) + \cdots + H_n(n) \tag{11}$$

with the consequent inequalities

$$H_B \geqslant H_k(k) \tag{12}$$

for each $k = 2, 3, \ldots, n$. The symmetry of all the f_s makes it unnecessary to indicate which particle is singled out in each of the terms of (11) or (12).

B. Consequences of Increasing H_B

A system may evolve irreversibly toward states characterized by greater values of H_B; ideally the evolution should be consistent with the n-body equations of motion. But the utility of something like the H-theorem is precisely that it permits statements to be made about the kinds of evolution that will occur without need to solve the equations of motion. However, it is clear that evolution that results in a changed value of H_G

has violated the equations of motion. Constancy of H_G is necessary, but not sufficient, to assure that the evolution has been consistent with the equations of motion Thus a weaker constraint might be applied: states are sought with greater values of H_B, but the states must have the same values for the first integrals (particle number, centroid momentum, angular momentum, total energy) and they must have the initial value of H_G. A variational calculation might be formulated to find the state of maximum H_B subject to these constraints. This is the usual way of using the H-theorem, but in this case one extra constraint has been imposed.

The constrained problem implies evolution in which the inequalities (12) become stronger and stronger, as a consequence of (11). But these inequalities require that the conditional probabilities (9) tell much more about particle positions than does the single-particle distribution, f_1. Larger values of H_B (under the constraint on H_G) indicate states of stronger particle correlation. The particle correlation does not mean that particles are necessarily close together; rather it implies that knowledge of the phase of one particle tells something about where other particles are to be found. Even if there is no upper bound to values of H_B (there is none for the stellar dynamical problem; an infinite reservoir of internal energy is available), evolution should proceed toward states of stronger correlation, and away from uncorrelated states like those described by (8). If a maximum for H_B can be attained, it is reached by maximizing correlations.

This argument may be illustrated through the simple example of a bivariate normal frequency function

$$f_2(x, y) = (2\pi\sigma^2)^{-1}(1 - \rho^2)^{-1/2} \exp\left[-\frac{x^2 + y^2 - 2\rho xy}{2\sigma^2(1 - \rho^2)}\right] \quad (13a)$$

which projects to the marginal frequency

$$f_1(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{x^2}{2\sigma^2}\right] \quad (14a)$$

but has conditional probability frequency

$$c_2(y \mid x) = (2\pi\sigma^2)^{-1/2}(1 - \rho^2)^{-1/2} \exp\left[-\frac{(y - \rho x)^2}{2\sigma^2(1 - \rho^2)}\right] \quad (15a)$$

The frequency $f_2(x, y)$ is constant on ellipses whose principal axes are at $\pm 45°$ in the x-y plane. When ρ is near ± 1 (strong correlation), the ellipse is elongated along the $\pm 45°$ lines. The marginal frequency, $f_1(x)$, is the projection of the ellipse onto the x-axis; it is a Gaussian characterized by σ. The conditional frequency, $c_2(y \mid x)$, is the frequency along a cut through the ellipse parallel to the y-axis; it is a Gaussian characterized by

$\sigma\sqrt{(1-\rho^2)}$, and is much narrower (for ρ near 1) than the marginal f_1.

The H-functions associated with these frequencies are

$$H_G = H(x, y) = \log(2\pi e) + \log(\sigma^2\sqrt{1-\rho^2}) \tag{13b}$$

$$H_B = H(x) = \log(2\pi e)^{1/2} + \log(\sigma) \tag{14b}$$

and

$$H_2(y) = \log(2\pi e)^{1/2} + \log(\sigma\sqrt{1-\rho^2}) \tag{15b}$$

These can easily be seen to obey (11) and the inequality (12). The Liouville theorem requires that the area inside an ellipse along which $f_2 =$ constant be preserved, and thus that the product $\mathscr{A}^2 = \sigma^2(1-\rho^2)$ (the determinant of the matrix of the quadratic form) be constant. The single-particle (marginal, or "Boltzmann") H_B is maximal if $\rho \to \pm 1$, and

$$H_B = \log(2\pi e)^{1/2} + \log(\mathscr{A}) - \tfrac{1}{2}\log(1-\rho^2) \tag{16}$$

can become arbitrarily large. The ellipse becomes arbitrarily long (and narrow); if $\rho \to +1$, the conditional probability shows that y is strongly localized near $y = x$; with $\rho \to -1$, y is just as strongly localized near $y = -x$. This last case illustrates a situation in which two particles (if the values of x and of y are regarded as particle coordinates for a two-particle one-dimensional system) need not be close together, but in which knowledge of where one particle is (x) tells where to look to find the other particle (y).

This example may easily be generalized to multivariate Gaussians: in that case, if M^{-1} be the matrix of the quadratic form which is the argument of the Gaussian, constancy of H_G is equivalent to demanding that the determinant of M be constant, while H_B is maximized by making the diagonal elements of M as large as possible. Again, H_B can become arbitrarily large by causing M to become ill-conditioned, so it has large diagonal elements for a given value of the determinant. The ellipsoid along which the frequency function is constant becomes very long and narrow (needlelike); its principal axes become quite unequal. But this is precisely the condition of strong correlation. The multivariate Gaussian describes spatially uniform Maxwellian systems. The mean square momentum increases as H_B increases, so the kinetic energy increases as well. The process stops when as much internal energy as possible has been converted into kinetic energy.

C. Equilibrium of Stellar Systems

Thermodynamic systems reach an equilibrium when the entropy is maximum (thus presumably H_B is a maximum, but we maintain a clear distinction between H_B and the thermodynamic entropy) *and* the internal

energy is a minimum. The minimum value for the internal (potential) energy implies an upper bound to the total kinetic energy, and thus to H_B for such systems, particularly if they attain a Maxwellian equilibrium.

Equilibrium cannot be defined for a stellar system by those rules. With stellar dynamical systems, there is no minimum to the internal (potential) energy; a stellar system can conserve total energy and still reach an infinite kinetic energy by calling on its infinite reserve of potential energy. (Real stellar systems are so far away from any limitation due to the finite size of stars that extremely drastic changes would appear long before the limit could be reached.) The only bound to admissible values of H_B is set by (7); H_B may become arbitrarily large.

The ensemble-averaged potential energy is given by (all masses are taken equal; the stars are treated as indistinguishable)

$$\begin{aligned}
\mathscr{V} &= -Gm^2 \frac{n(n-1)}{2} \\
&\quad \times \int d^{3n}x\, d^{3n}v \frac{f_N(x^{(1)}, x^{(2)}, \ldots, x^{(n)}; v^{(1)}, v^{(2)}, \ldots, v^{(n)}; t)}{|x^{(\alpha)} - x^{(\beta)}|} \\
&= -Gm^2 \frac{n(n-1)}{2} \\
&\quad \times \int d^3x^{(\alpha)}\, d^3x^{(\beta)} \frac{1}{|x^{(\alpha)} - x^{(\beta)}|} \int d^3v^{(\alpha)}\, d^3v^{(\beta)} f_2(\alpha, \beta; t) \\
&= -Gm^2 \frac{n(n-1)}{2} \int d^3x^{(\alpha)}\, d^3x^{(\beta)} \frac{n_2(x^{(\alpha)}, x^{(\beta)}; t)}{|x^{(\alpha)} - x^{(\beta)}|} \qquad (17)
\end{aligned}$$

where $n_2(x^{(\beta)}, x^{(\alpha)}; t)$ is the projection of f_2 onto the six-dimensional pair configuration space. Chandrasekhar has repeatedly stressed the dependence of the potential energy on the pair configuration distributions, n_2.[21] When the pair correlations are introduced as an additive part in the usual manner,

$$f_2(1, 2) = f_1(1)f_1(2) + g(1, 2) \qquad (18)$$

the total potential energy breaks up into the sum of two pieces. There is the usual part obtained from f_1's

$$\mathscr{V}_1 = -Gm^2 \frac{n(n-1)}{2} \int d^3x^{(1)} n(x^{(1)}) \int d^3x^{(2)} \frac{n(x^{(2)})}{|x^{(2)} - x^{(1)}|} \qquad (19)$$

and a second part due to the pair correlation:

$$\mathscr{V}_c = -Gm^2 \frac{n(n-1)}{2} \int d^3x^{(1)}\, d^3x^{(2)} \frac{1}{|x^{(2)} - x^{(1)}|} \int d^3v^{(1)}\, d^3v^{(2)} g(1, 2) \qquad (20)$$

This is not an idle exercise; real stellar systems show appreciable pair correlation, just as do n-body systems in a computer (Section III). The contribution of $\mathscr{V}_c$ to the total energy can become quite large. The easiest way to get large contributions for $\mathscr{V}_c$ is to form binaries, to place near each star one other in a tightly bound Keplerian orbit. This can be done such that f_1, and hence $\mathscr{V}_1$, do not change, while obeying the other constraints (except that in H_G). Indeed, if an "equilibrium" ratio of binary to single systems is determined by the usual Boltzmann factors ($\exp -\chi/kT$), all the stars want to be members of binary systems [12]!

There have been attempts[9–11] to discuss equilibrium configurations for stellar systems by means of a variational calculation in which f_1 is varied to maximize H_B under constraints of constant particle number and constant total energy. These are usually done in a framework in which the constraint of constant H_G is not applied, and in which $\mathscr{V}_1$ is the only part of the potential energy accounted for in the energy bookkeeping. Treatments in which $\mathscr{V}_c$ is ignored seem implicitly to take the n-particle distribution in the form of a product of f_1's as in (8); if so, $nH_B = H_G$, and H_B is a constant of the motion. The postulated evolution takes place in violation of the Liouville theorem and hence of the n-body equations of motion. An alternative viewpoint is that the inequalities (12) show that increases of H_B must be accompanied by growth of correlation in the n-body system. To ignore $\mathscr{V}_c$ is to ignore an essential feature of the evolution; it is at best inconsistent.

A technical point remains: these assertions are unfortunately not rigorous. It is possible to construct f_N's with all the required symmetries and such that $f_2(1, 2) = f_1(1) \cdot f_1(2)$ (i.e., such that $g(1, 2) = 0$), but which still show correlation (at higher orders) and hence would satisfy the inequalities (12) with $\mathscr{V}_c = 0$. However, the only examples known to the writer are rather special—so special that it can safely be asserted that an essential part of a calculation in which $\mathscr{V}_c$ is taken to be zero is a demonstration that f_N actually has this property. Similarly, it is possible to construct functions which, inserted for g in (20), yield $\mathscr{V}_c = 0$ (by a Gram-Schmidt process, for example); but it is not clear that a function f_N can be found for which these g's can be generated by (5) and (18). Again, the functions are rather special, and it would seem to be an essential part of any calculation to show that $\mathscr{V}_c = 0$ for the f_N used. In general, this is not a simple program.

Indications that correlations are important in self-gravitating systems are not new. Prigogine and Severne[7] constructed a weak-coupling kinetic theory of binary interactions which predicted an irreversible growth of correlational energy. Their model was spatially uniform and of infinite extent, and the arguments did not hinge on H_B.

D. Rate of Increase of H_B

The rate at which systems evolve toward states of greater H_B is important for application to stellar systems. Doremus, Feix, and Baumann[22] have shown that spherically symmetrical systems described by a Vlasov equation with a distribution that depends on the energy alone are stable if $f_1(E)$ is monotone decreasing with increasing energy, E. These conditions can be met easily for realistic model systems. When the Liouville equation is projected from Γ to μ there is a term

$$-\int d^3x^{(2)}\, d^3v^{(2)} \frac{\partial}{\partial x^{(1)}}\left(\frac{1}{|x^{(2)} - x^{(1)}|}\right)\frac{\partial}{\partial v^{(1)}} g \tag{21}$$

on the right-hand side; taking that term to be zero gives the Vlasov equation. Since systems without the term (21) are stable, the evolution described by increasing H_B must proceed through that term. This term describes the effect of binary encounters; in normal stellar problems, the time scale indicated is that of the usual relaxation time that is calculated from binary encounters. For most systems, that time scale is quite long; it is a secular, and not a dynamical, time scale. Larson[23] has confirmed that changes in the direction of increasing H_B proceed on a relaxation time scale. His formalism is quite different from that used here, and does not allow for pairing effects. Evolution of f_1 involves secular changes in f_2. The equation governing the change in f_2 involves the three-particle correlation; but since the correlation terms of all orders are of order unity (in the particle number; Section III), that change may be expected to proceed no faster than the relaxation time scale. This argument about the rate presupposes obedience to the equations of motion; this is the only premise available to estimate a rate.

E. Summary

If actual stellar systems evolved toward increased H_B, they would undergo a secular trend toward states characterized by a preponderance of binaries, since this is the easiest way to increase correlation. There is no observational evidence that cluster evolution proceeds in this way. Tight, dynamically formed binaries are indicated by this argument—the observed binaries are thought to have been formed as binaries, and not to indicate the kind of evolution implied by increasing H_B. The fact that very few, if any, of the observed binaries are thought to have been dynamically formed seems to imply some mechanism that inhibits binary formation. Such a mechanism is also indicated by n-body numerical experiments.[18,19] Lynden-Bell[12] has also noted the need for a mechanism to inhibit the

dynamical formation of binaries. He concludes that this mechanism somehow permits the pairing term to be ignored in a variational calculation designed to maximize H_B; that does not seem to be a satisfactory way out of the dilemma posed by the H-theorem's demand for correlation. The mechanism is discussed from another point of view in Section V.

Evolution toward greater values of H_B also runs against the virial theorem. As correlation increases, the magnitude of the potential energy increases; the total kinetic energy, K, must also increase in order to conserve total energy. The ratio $-\mathcal{V}/K$ must approach 1, not the 2 of the virial theorem. Both observational data and n-body calculations seem to confirm the 2 of the virial theorem over the 1 of the variational calculation.

Evolution toward greater values of H_B does not seem to provide a useful technique for studying stellar systems on two grounds: because H_B can become infinite without local maxima, it does not predict stable equilibria; and the evolution predicted is not in agreement with observation. It also does not predict metastable equilibria, but this requires arguments not presented in this paper.[24] At any rate, such evolution, far from being a "catastrophe," must proceed on a secular (relaxation) time scale. Note that nothing has been said about those versions of the H-theorem that make use of coarse-graining.

V. THE MICROCANONICAL ENSEMBLE

The microcanonical ensemble is the only ensemble appropriate to stellar systems, which are presumed to obey the n-body equations of motion and to conserve total energy. The canonical and grand canonical ensembles are not suitable because the notion of a heat bath is alien to stellar systems and because the long-range forces preclude treatment of subdivisions of the total system. The difficulties with numerical experiments, described in Section II, underline these features. Although this seems to prevent discussion of "isolated" stellar systems, one liberty that is taken is to assume that the dynamics of a cluster can be discussed without specifically taking the other known clusters into account. This is a reasonable restriction because near neighbors among clusters or among galaxies are usually much farther away than the typical dimensions of the cluster under consideration. In part, the rationalization for this simplification is one of necessity; even so, the problem is intractable. If the dynamics of such a system could then be worked out, however, the next step would be to study its stability. Given a stable system, the effects of neighbors could be introduced as perturbations.

The microcanonical ensemble is a useful illustration, and brings out some unexpected features of stellar dynamical systems. We begin with the

evaluation of the volume of phase space under the energy surface, $E =$ constant ($E < 0$). The topology of energy surfaces is sufficiently smooth to permit a straightforward attack on the problem.

Let θ be the step-function. Then

$$\Omega(E) = \int d^{3n}x_i^{(\alpha)} \int d^{3n}p_i^{(\alpha)} \delta\left(\sum_\alpha m_\alpha x_i^{(\alpha)}\right) \delta\left(\sum_\alpha p_i^{(\alpha)}\right) \theta\left(E - \sum_{i,\alpha} \frac{(p_i^{(\alpha)})^2}{2m_\alpha} - \mathscr{V}\right) \tag{22}$$

where the delta functions express the invariance of the equations of motion to translation of the origin of coordinates, the p's are momenta, and E is the total energy. A delta function expressing the conservation of total angular momentum is omitted to simplify the discussion; its inclusion causes no qualitative changes in the situation. Greek subscripts or superscripts denote particles, $\alpha = 1, 2, \ldots, N$; and Latin letters are used for the usual vector subscripts, $i = 1, 2, 3$. The summation convention is not used. All masses will be taken to be the same, so the subscript on m is unnecessary.

The integral over momenta can be carried out immediately by recognizing that the step function selects the interior of a $3N$-dimensional sphere of radius $[2m(E - \mathscr{V})]^{1/2}$. The three delta functions in p reduce the dimensions by three: when a sphere is cut by a plane through its center, the result is a sphere of one less dimension. Thus, apart from the numerical coefficient $A = (2m\pi)^{(3N-3)/2}/\Gamma[(3N - 1)/2]$, the momentum integration yields the result

$$\Omega(E) = A \int d^{3N}x_i^{(\alpha)} \delta\left(\sum_\alpha m x_i^{(\alpha)}\right) (E - \mathscr{V})^{(3N-3)/2} \tag{23}$$

This integral may easily be seen to be of the form

$$\Omega(E, R) = C_N(R) \int_{-E}^{\infty} d(-\mathscr{V})(-\mathscr{V})^{2-3N}(E - \mathscr{V})^{(3N-3)/2} \tag{24}$$

since the potential energy is a homogeneous form of order -1 in the coordinates. The quantity $(Gm^2/(-\mathscr{V}))$ plays the role of the radius in a $3N$-dimensional polar coordinate system. Alternatively, a coordinate system can be constructed in which $(-\mathscr{V})$ is one of the coordinates; the Jacobian of the transformation to this new coordinate system is a homogeneous function in $(-\mathscr{V})$ of order -2 times its rank, and the same result follows. Because explicit forms of the new coordinate systems are of use only to show that the exercise can be carried through, there is no need to display the construction.

The dimensional argument leaves the numerical coefficient undetermined. The coefficient can be obtained from A and the configuration

volume between two surfaces $(-\mathscr{V}) =$ constant separated by $d(-\mathscr{V})$. It is easy to see that the surfaces are well enough behaved that the volume inside a surface $(-\mathscr{V}) =$ constant can be calculated for a finite part of the configuration space; the required coefficient can be obtained by differentiating that volume. But the volume inside a surface $(-\mathscr{V}) =$ constant can be infinite because an arbitrarily large amount of potential energy is available by letting two particles come close together; all the remaining particles can then be removed to infinity. All but three particles can be removed to large distance, but the volume inside $(-\mathscr{V}) =$ constant is infinite for three particles by leaving two close together and removing the third, and so on. The contributions of all these combinations may readily be summed. The growing infinity must soon overtake any finite contribution from interesting parts of the $(-\mathscr{V})$ surfaces near the origin; thus for large enough distances (R large) the volume associated with $d(-\mathscr{V})$ tends asymptotically to

$$C_N(R) = \frac{(2m\pi)^{(3N-3)/2}}{\Gamma[(3N-1)/2]} \frac{N(N-1)}{2} 4\pi(Gm^2)^{3N-3} \frac{\pi^{3(N-2)/2}}{\Gamma[(3N-4)/2]} (2R^3)^{N-2} \tag{25}$$

The limiting volume is a sphere in the $(3N-3)$-dimensional configuration space that is left after the delta functions that keep the centroid on the origin have been incorporated. This provides a convenient cutoff because it, too, is a homogeneous form in all the coordinates and thus avoids introducing some awkward corners in which the potential cannot be treated as a homogeneous function. The term R^3 can be simply expressed as a three-dimensional volume, V.

The divergence of $\Omega(E, R)$ is associated wholly with the configuration integral: the momentum integral leads to a finite result. The large negative power on $(-\mathscr{V})$ is more than adequate to suppress the divergence on the $(E - \mathscr{V})$ term as $(-\mathscr{V})$ becomes infinite.

The cutoff is unpleasant, but it might be placed well outside the present position of any particles, and can be imagined to remain outside of all particles. It is introduced for formal purposes, and can be treated as cutoffs usually are in those unfortunate cases where they are required. Most calculations can be formulated with limits to be taken as $R \to \infty$; it can sometimes be arranged so that the dependence on $C_N(R)$ cancels before the limit is taken.

According to (24), the phase volume inside an energy surface is

$$\Omega(E, R) = C_N(R) B\left(\frac{3N-3}{2}, \frac{3N-1}{2}\right)(-E)^{(3-3N)/2} \tag{26}$$

and the surface area (the phase volume available at constant energy within some ΔE) is

$$\sigma(E, R) = \frac{\partial \Omega}{\partial E} = \frac{3N - 3}{2} C_N(R)$$

$$\times \int_{-E}^{\infty} d(-\mathcal{V})(-\mathcal{V})^{2-3N}(E - \mathcal{V})^{(3N-5)/2}$$

$$= (2m\pi)^{(3N-3)/2} \frac{(Gm^2)}{(3N - 3)!} \frac{2^{N-1}N(N - 1)}{\sqrt{\pi}}$$

$$\times \frac{\Gamma[(3N - 1)/2]}{\Gamma[(3N - 4)/2]} (R^3)^{N-2}(-E)^{(1-3N)/2} \tag{27}$$

The function $B(c_1, c_2)$ is the beta function. The integrals diverge strongly for non-negative total energy, $E \geqslant 0$.

A. Meaning for the *n*-Body Problem

The meaningful case for the n-body problem is that for $E =$ constant, on the energy surface as in (27). The integrand of (27) may be regarded as a probability distribution function for $(-\mathcal{V})$. More negative values of $\mathcal{V}$ lead to rapidly growing phase volumes in the momentum space through the $(E - \mathcal{V})^{(3N-5)/2}$ term. Since $(-\mathcal{V})$ may become infinite by letting two or more particles come close together in the configuration space, an infinite volume (at quite a high order infinity!) opens up in the momentum space. The discussion is usually terminated at this point with the observation that there is infinite phase volume because of close binaries. However, the configuration volume available for such particle aggregates diminishes faster than the momentum space opens up, leaving a net decrease in the available phase volume. The most probable values for $(-\mathcal{V})$ are those in which the two terms just play off against each other.

This casts the problem of dynamical formation of binaries in quite a different light: binaries do not form because there is *not* a preponderance of phase volume accessible for them. The same argument applies to the formation of higher-particle aggregates as well. This feature is in accord with the lack of binaries thought to be dynamically formed among the observed binaries, and may well be the mechanism inferred in Section IV.D. This does not mean that binaries do not form—it means that these processes are not strongly preferred in terms of the available phase space.

The integral of (27), interpreted as a probability distribution for $(-\mathcal{V})$, leads to an interesting observation. The ensemble-averaged expectation

for $(-\mathscr{V})$ is just twice $(-E)$—the virial theorem:

$$\langle -\mathscr{V} \rangle = \lim_{R\to\infty} \left\{ \frac{C_N(R) \int d(-\mathscr{V})(-\mathscr{V})(-\mathscr{V})^{2-3N}(E-\mathscr{V})^{(3N-5)/2}}{C_N(R) \int d(-\mathscr{V})(-\mathscr{V})^{2-3N}(E-\mathscr{V})^{(3N-5)/2}} \right\}$$

$$= 2(-E) \tag{28}$$

This is a manifestation of ergodicity; the ensemble average and the time average lead to the same result. The probability distribution interpretation of (27) could possibly be used in discussions of mass determination for clusters of galaxies or of star clusters by the virial theorem (see Ref. 25 for some recent papers on this subject). An especially valuable consideration is that this formulation is not based on assumptions of equilibrium or of the absence of interactions. The variance estimated for the ratio that appears in the virial theorem sharpens up appreciably when there are more particles in the system:

$$\text{Ave}\left(\frac{-\mathscr{V}}{-E}\right) = 2\ ; \qquad \text{Var}\left(\frac{-\mathscr{V}}{-E}\right) = \frac{4}{3N-5} \tag{29}$$

Subject to rather poor statistics, these and other higher moments of this distribution were confirmed by means of n-body calculations some years ago. With E positive, the expectation for $\mathscr{V}$ is zero, as might be expected.

This result concerning the n-body problem also tallies with one of the observed features of n-body calculations. The comment that the term $(E-\mathscr{V})^{(3N-5)/2}$ leads to infinite momentum space for the formation of binaries suggests that the binaries formed in n-body calculations might be a manifestation of numerical error,[14] which causes the computed system to diffuse about over the integral hypersurface; it should preferentially diffuse toward the formation of more binaries because there is much more phase space available there. If that were the explanation, n-body calculations should not yield good values for the virial ratio $(-\mathscr{V})/K$; but they do. If many tight binaries were to form because that is a preferred state in terms of the available phase space, the virial theorem should be violated: both the total kinetic energy (K) and the total potential energy $(-\mathscr{V})$ should increase, and the ratio should approach one, not the observed value of two. The binaries reported by Aarseth,[26] van Albada,[27] and others cannot simply be ascribed to numerical error because the computed system does not show the corollary effects on the virial ratio that should follow if the cause is numerical error. For computer n-body calculations to be as stable as they are in the fact of very severe problems of numerical

accuracy[14] strongly argues in favor of the interpretation that the available phase volume is given by an expression like (27).

B. "Thermodynamics"

The area of the energy hypersurface can be used to key into the usual thermodynamic analogies of statistical mechanics through the relation $\exp(S) = \sigma(E)$,

$$S = \text{const.} + (N - 2)\log(\mathbf{V}) - \frac{3N - 1}{2}\log(-E) \tag{30}$$

where $\mathbf{V}$ is the volume inside the cutoff distance, R, that appears in (25). The constant in this equation is distinct from that always present in an entropy in classical statistical mechanics because there is no natural volume element in the phase space; but the two may be lumped together with no harm. In view of the warnings of Landau and Lifshitz[28] about the inapplicability of the second law to problems involving gravitation, this should perhaps be regarded as an exercise. Furthermore, the preceding discussion concerns the system as a whole, and does not allow subdivision of the system. Most thermodynamic arguments presuppose that a small portion of the system of interest can be isolated and studies by itself, without changing the properties by the very process of isolating the subsystem. That is not the case with self-gravitating systems. Nonetheless, some interesting features arise, and the formulation that results from blind application of the cookbook rules seems internally consistent if some care is exercised in the interpretation of the results.

The usual thermodynamic definitions of temperature and specific heat lead to results that are qualitatively well known:

$$\begin{aligned}\frac{1}{T} &= \frac{dS}{dE} \\ &= \lim_{\mathbf{V}\to\infty} \frac{d}{dE}\left[\text{const.} + (N - 2)\log\mathbf{V} - \frac{3N - 1}{2}\log(-E)\right] \\ &= \frac{3N - 1}{2(-E)}\end{aligned} \tag{31}$$

$$C_V = \left(\frac{\partial E}{\partial T}\right)_{\mathbf{V}} = -\frac{3N - 1}{2} \tag{32}$$

The temperature, which is positive, is defined in terms of the total energy, E, and not in terms of the kinetic energy per particle, as sometimes written. There is a (probabilistic) relation between the two through the virial theorem, but that is not as sharp as might be desired. The specific heat at constant volume is negative. The virial theorem and the negative specific

heat both follow from the formulation. These analogies cannot be pushed too far; the entropy does not go to zero as the temperature goes to zero, so there is no Nernst's theorem unless things are carefully arranged to produce a cancellation. Similarly, the specific heat does not go to zero at zero temperature.

The entropy contains a term due to the cutoff, which is written as a volume in (30). Formally, this gives rise to a pressure (which is due to particles that reach the cutoff distance in violation of the spirit of a cutoff):

$$P = -\left(\frac{\partial E}{\partial \mathbf{V}}\right)_S = \text{const. } \mathbf{V}^{-(N+3)/(3N-1)} \tag{33}$$

Again, if the limit $\mathbf{V} \to \infty$ is taken, the pressure goes to zero at least as fast as $1/R$. (However, the product $P\mathbf{V}$, which appears in the thermodynamic forms of the virial theorem, diverges.) This should not be interpreted as a polytropic relation, notwithstanding the power-law dependence. This relation holds for an isolated large system, in terms of dimensions substantially larger than the system itself, and is not the kind of local dependence of pressure on density for an isolated part of a larger system that is implied by a polytropic equation of state. Although more of the standard thermodynamic functions might be evaluated, that does not seem to be a fruitful undertaking.

C. Equilibrium of Star Clusters

The solution for the microcanonical ensemble presented here suggests that the only equilibrium for star clusters is a state with a single close binary, with all other particles at infinity. This has long been a formal difficulty with the famous "three-body problem": the solutions in which all three particles are near the origin are a set of measure zero relative to the measure of the available phase volume, which has the same infinity shown here for more particles. Although aggregates of more than two particles can lead to infinite contributions as well, the infinity associated with each m-particle aggregate is dominated by two-particle term with the remaining $(m - 2)$ particles well separated from the pair and from each other. The two-particle aggregates dominate because m-particle aggregates lead to coefficients like $C_N(R)$ that only have R to the lower power, $R^{3(N-m)}$.

In the study of star clusters, we do not want a true ensemble average or (assuming ergodicity) a true time average: either of these would only tell about the state with a single binary and infinite dispersion of the other particles. The desired averages are over systems with most of the particles near the origin.

This tendency toward dissolution of the clusters is probably associated with the infinite phase volume available if the total energy is positive. The total energy can be made positive for a subgroup of particles by absorbing all the negative energy into a single pair of particles. This conjecture about the tendency toward dissolution might be proved by the use of the reduced distribution functions obtained from the microcanonical ensemble; but this is not a straightforward program because the integrals done so far have been worked out by appeal to the fact that the potential energy is a homogeneous function of the configuration coordinates. Once one or more coordinates are fixed, as they must be to evaluate a reduced distribution, the potential is no longer homogeneous and a new technique must be developed to evaluate the integrals.

Similarly, the rate of dissolution might be worked out by computing the flux of the distribution across a boundary in configuration space. Again, the placement of such a boundary vitiates the trick used to compute the integrals. It is believed, on the other grounds, that the problem of evaporation of stars from clusters is dominated by the secular, or "relaxation" time scale[1]; this is the very long time scale referred to earlier. Clusterings into aggregates of particles appear to be quite long-lived metastable states which are totally lost in the crude view of cluster dynamics provided by the microcanonical ensemble.

D. The H-Theorem

The development of the microcanonical ensemble does not proceed in the direction that would cause the most rapid increase of H_B. The argument is based on the nature of the infinity in the phase volume. Larger values of H_B imply that there are conditional probabilities such that some knowledge of the state of the system allows more precise statements to be made about the total state than could be made in the absence of that knowledge. But the final state of the microcanonical ensemble is the antithesis of that condition: the preponderance of phase volume is dominated by states such that knowledge of the phase of one particle only carries information $O(1/n)$ about which particle should be near some other particle. While H_B might become arbitrarily large with correlations $O(1/n)$, it could become much larger still with correlations of order unity (in the particle number).

It is difficult to make this argument precise because of the various infinities lurking in the background. However, it is possible to construct ensembles for the gravitational n-body problem that yield finite H_G, and for which, therefore, the subsequent evolution is not in directions that make H_B increase as rapidly as possible. This may be done as follows.

The phase volume, $\sigma(E)\,\Delta E$, computed according to (27), corresponds to minimal information about the system: an ensemble is constructed that contains examples of all systems having the same number of particles and the same total energy. The ensemble should fill the volume $\sigma(E)\Delta E$ with uniform density, and have $f_N = 0$ elsewhere. Imagine the system with a finite value of the cutoff distance, R, so the total volume may be finite. If H_G is computed for this ensemble, the value obtained is just S of (30). This is the greatest H_G consistent with the given value of the first integrals and the restriction in volume. The extra information that the system is known (at some time) to be in a finite part of the configuration space is sufficient to assure a finite value for H_G.

The microcanonical ensemble does not provide a valid counterexample to the H-theorem. The evolution from an initial state of finite H_G, and possibly with $nH_B = H_G$, is toward a state of greater H_B. There are no indications of systematic trends in which H_B decreases over any reasonable interval of time. However, the evolution is not of a kind that would make H_B increase as rapidly as possible because the final correlation is only $O(1/n)$. Somehow, the H-theorem would seem to carry, as a corollary, the implication that evolution should proceed in directions that would make H_B increase as rapidly as possible—the microcanonical ensemble is a counterexample to that much weaker statement. At any rate, the H-theorem is not useful if there is no maximum to H_B (so that no terminal equilibrium state can be predicted) and the evolution does not proceed in such a way that the state of the system can be correctly predicted at later times by solving for a maximum of H_B while H_B is still finite.

VI. CONCLUDING REMARKS

The numerical experiments and discussions of the preceding sections illustrate that considerable care is required in attempts to use the methods of statistical mechanics with stellar dynamical problems. Stellar systems behave in qualitatively different ways from the systems usually dealt with in statistical mechanics. But this should have been expected; the observed tendency of stellar systems to cluster or to form multiparticle aggregates is one manifestation of this difference. Stellar systems have a long-time memory of initial conditions; the equations of motion and the boundary conditions do not admit the usual mechanisms that make statistical methods work. Elaboration of these points leads to matters already mentioned in the introduction, with stress on finite systems, the nature of collisions, non-Markovian character, and so on.

Correlations to all orders are important; they are all of order unity in the particle number. Because the pair terms can lead to a large contribution to the potential energy, correlation effects must be taken into

account in the energy bookkeeping. A new series of numerical experiments (different from those of Section III) is under way to study correlation energy. The experiments are still at a preliminary stage, but seem to show that the (time-averaged) correlation energy settles down at a value near the total cluster energy. About half of the (time-averaged) potential energy in the cluster is correlation energy. This indicates that $\mathscr{V}_c$ need not be small.

The H-theorem is not useful for stellar systems. This is not a proof that there is no H-theorem. The statement about usefulness is based on two points. First, H_B has no maximum (either local or global), so an H-theorem cannot be used to find equilibrium distributions. Second, the trends toward greater H_B shown by the one example worked out are not those that would make H_B increase most rapidly. In the absence of a maximum, an H-theorem might still be useful if the evolution of actual systems maximized the rate of increase of H_B; but an example has been displayed in which this is not so. Further, the trend that would make H_B increase most rapidly, that toward formation of binaries, is contrary to observation.

No equilibrium solution has been found by any of the methods used here. Some equilibrium solutions are known: among them various configurations of the three-body problem, n particles arranged at the vertices of a regular polygon and rotating in their plane, and several rigidly rotating planar configurations derived from the polygon. Most of these are unstable. At best these configurations represent a "set of measure zero" relative to all the possible configurations, so they might understandably escape detection by these methods. The criteria used to define equilibrium may be too strict; certainly it is not realistic to demand stability over times greatly in excess of the age of the universe. It seems likely that we are faced with a situation in which there may be no equilibrium in a mathematical sense, but in which the natural processes leading to dissolution of clusters are so slow that clusters represent an equilibrium in a practical sense. An alternative is that other processes (dissipation due to interstellar gas, etc.) might be more important than we have believed them to be, and that some other mechanism may be responsible for the apparent equilibria observed in nature. The challenging puzzle still stands.

References

1. S. Chandrasekhar, *Principles of Stellar Dynamics*, Dover, New York, 1960.
2. D. Lynden-Bell, *Monthly Notices Roy. Astron. Soc.*, **124**, 279 (1962); E. P. Lee, *Astrophys. J.*, **148**, 185 (1967).
3. D. Lynden-Bell, in *Relativity Theory and Astrophysics*, J. Ehlers, Ed., Vol. 2, *Galactic Structure*, American Mathematical Society, Providence, R.I., 1967, pp. 131–168.
4. J. P. Ostriker and A. F. Davidson, *Astrophys. J.*, **151**, 679 (1968).

5. S. Chandrasekhar and J. von Neumann, *Astrophys. J.*, **95,** 489 (1942); **97,** 1 (1943); S. Chandrasekhar, *Rev. Mod. Phys.*, **15,** 1 (1943), reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed., Dover, New York, 1954, pp. 3–91; S. Chandrasekhar, *Astrophys. J.*, **99,** 25 (1944); **99,** 47 (1944).
6. E. P. Lee, *Astrophys. J.*, **151,** 687 (1968).
7. I. Prigogine, *Nature*, **209,** 602 (1966); I. Prigogine and G. Severne, *Physica*, **32,** 1376 (1966); I. Prigogine and G. Severne, *Bull. Astron.* 3 (3), 273 (1968).
8. R. Kurth, *Introduction to the Mechanics of Stellar Systems*, Pergamon, New York, London 1957.
9. V. A. Antonov, *Vestn. Leningr. Univ.*, **7,** 135 (1962). This paper is usually difficult to locate, but much of the argument is repeated in Ref. 10.
10. D. Lynden-Bell and Roger Wood, *Monthly Notices Roy. Astron. Soc.*, **138,** 495 (1968).
11. W. Thirring, *Z. Physik*, **235,** 339 (1970).
12. D. Lynden-Bell, "Statistical Mechanics of Stellar Systems," in M. Chretien, S. Deser, and J. Goldstein, Eds., *Astrophysics and General Relativity*, Summer School on Theoretical Physics at Brandeis University, 1968, Gordon and Breach, New York, 1969.
13. W. C. Saslaw, *Monthly Notices Roy. Astron. Soc.*, **141,** 1 (1968); **143,** 437 (1968); **147,** 253 (1970).
14. R. H. Miller, *J. Comp. Phys.*, **8,** 449 (1971).
15. Berni J. Alder, in *Proceedings of Symposia in Applied Mathematics*, Vol. XV, "Experimental Arithmetic, High Speed Computing, and Mathematics," American Mathematical Society, Providence, R.I., 1963, pp. 335–349.
16. S. von Hoerner, *Z. Astrophys.*, **50,** 184 (1960); **57,** 47 (1963).
17. S. J. Aarseth, *Monthly Notices Roy. Astron. Soc.*, **126,** 223 (1963); **132,** 35 (1966); *Astrophys. Space Sci.*, **14,** 118 (1971). R. Wielen, *Ver. Astron. Rechen-Instituts Heidelberg*, No. 19 (1967); *Bull. Astron.*, **3** (3), 127 (1968).
18. R. H. Miller, *Astrophys. J.*, **165,** 391 (1971).
19. R. H. Miller, *Astrophys. J.*, **172,** 685 (1972).
20. A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York, 1957; C. E. Shannon, *Bell System Tech. J.*, **27,** 379, 623 (1948), reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Ill., 1949.
21. S. Chandrasekhar and E. P. Lee, *Monthly Notices Roy. Astron. Soc.*, **139,** 135, (1968); S. Chandrasekhar and D. D. Elbert, *Monthly Notices Roy. Astron. Soc.*, **155,** 435, (1971).
22. J.-P. Doremus, M. R. Feix, and G. Baumann, *Phys. Rev. Lett.*, **26,** 725 (1971).
23. R. B. Larson, *Monthly Notices Roy. Astron. Soc.*, **147,** 323 (1970).
24. R. H. Miller, *Astrophys. J.*, **180,** 759(1973).
25. D. Sher, *Astrophys. J.*, **171,** 537 (1972); S. J. Aarseth and W. Saslaw, *Astrophys. J.*, **172,** 17 (1972).
26. S. J. Aarseth, *Astrophys. Space Sci.*, **13,** 324 (1971).
27. T. S. van Albada, *Bull. Astron. Inst. Neth.*, **19,** 479 (1968).
28. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 2nd. ed., Addison-Wesley, Reading, Mass. 1969, Sec. 8.

ENZYME CASCADES AND THEIR CONTROL IN BLOOD PLASMA

ELLIOTT W. MONTROLL

Institute for Fundamental Studies, Departments of Physics and Chemistry, University of Rochester, Rochester, New York

CONTENTS

I. Introductory Remarks on the Nature of Proteins and Enzyme Kinetics . 147
II. Description of Several Enzyme Cascades 151
A. Blood Coagulation 151
B. Fibrinolytic Mechanism 163
C. Complement and the Complement Cascade 164
D. The Renin-Angiotensin System 166
E. The Kinin System 167
III. Kinetics of Enzyme Cascades 169
Acknowledgment 174
References 174

Since the uninterrupted, well regulated flow of blood is vital to the life of higher animals it is essential that mechanisms exist to respond to fluctuations induced by accidents and external agents. In the course of the evolution of those species which have survived, no outside expert was available to make quick repairs when fluctuations became too large. Hence the appearance of an assortment of molecules capable of inducing the chemical reactions and participating in the physical mechanisms basic to the spontaneous repair processes had to be concurrent with the development of the complex animals.

Any leaks that appear in the blood-conducting tubes have to be plugged before too much of the valuable fluid is lost. Invading organisms that penetrate into the bloodstream must be destroyed before their population multiples to a level that endangers the life of the host. The blood pressure must be controlled and the system should be capable of adjusting to the variations in the elasticity and the cross-section of the tubes of the circulatory system which result from aging.

The response mechanisms should be well tuned so that they are not activated by false signals. Spontaneous blood coagulation in arteries

without leaks is as dangerous as no coagulation to seal lesions. It seems likely that the large number of steps which precede the activation of the final repair process in enzyme cascades has evolved to distinguish signal from noise. If each of the preliminary steps were to follow some clue from the agent which demands the completion of the cascade, their large number would require a precise characterization of the agent. Then the probability that the normal harmless fluctuations in the system would mimic the agent in more than a small number of ways at a given time would be very small. The development strategy of such a control system would be analogous to the design strategy of a sophisticated sea mine which is constructed to explode under a special kind of ship such as an aircraft carrier and which is also supposed to be difficult for the enemy to sweep. A minesweeper with a noisemaker which simulates the noise of the carrier would activate only one component of the response system; a long electrical conductor towed by the sweeper might mimic the magnetic field of the hull of the carrier and activate the next component. However, unless the sweeper could also generate the special pressure field developed by such a long streamlined object in rapid motion in the water, the mine would eventually return to its inactive state without exploding.

The volume of clinical and biochemical literature on enzyme cascades, especially blood coagulation, probably exceeds that on almost any other specialized technical subject. Some of the important basic ideas have resulted from curious early observations which one would hardly expect to have had the same status as those from carefully planned experiments. Clever mechanisms still of interest were proposed long before there was any idea of the structure of the molecules involved. There is even a certain charm in finding important process factors still named after patients who lacked them—for example, Christmas, Hageman, and Stuart—rather than after the physicians who discovered them. When the Mendelian laws of heredity became generally known at the turn of the century one of the first applications to human genetics was the analysis of the inheritance of hemophilia. This was of course facilitated by known records of victims of the disease in royal families whose geneologies were well documented.

The process of hemostasis and the formation of the hemostatic plug involves (*a*) highly specialized cells of the blood stream, the platelets, which have an affinity for sites in the blood vessel where damage has occurred; (*b*) blood coagulation, the gel transition of monomer fibrinogen into a polymeric fibrin network in and around the aggregated platelets, stabilizing the plug; and (*c*) the dynamics of the blood vessel which leads to these processes and which finally must perform the healing process.

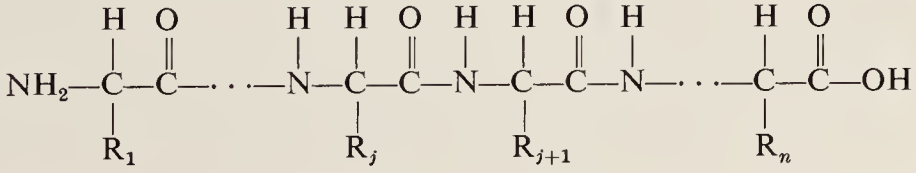
An important part of the extensive literature on hemostasis is to be found in the proceedings of many conferences on the subject. Some of the

conferences, such as the Hamburger Symposion über Blutgerinnung and the Wayne State Conferences, are held at regular intervals, as are those of various international committees on blood such as the Committee on Blood Clotting Factors and the Committee on Haemostasia and Thrombosis. One of the main journals on the subject, *Thrombosis et Diathesis Haemorrhagica*, has a supplement which contains the proceedings of many of the important conferences. The proceedings of the conference of the British Royal Society (organized by R. G. MacFarlane) on triggered enzyme systems in blood plasma (*Proc. Roy. Soc. B*, July 1, 1969) has been very valuable to the author in preparing this review.

Since the main intent of the review is to present a guide to the literature to chemists, physicists, and engineers who wish to learn something of the subject class of biological control mechanisms, it will be made more self-contained by starting with a few remarks on the nature of proteins and classical enzyme kinetics. An attempt has been made to use a minimum of medical terminology. If some has crept in with insufficient definition the uninitiated can find good explanations in one of the standard books such as Best and Taylor, *Physiological Basis of Medical Practice* (Williams and Wilkins). Two excellent books have recently appeared on blood clotting, one edited by W. H. Seegers and entitled *Blood Clotting Enzymology* (Academic Press, 1967) and the other, edited by K. Laki, entitled *Fibrinogen* (Dekker, 1968).

I. INTRODUCTORY REMARKS ON THE NATURE OF PROTEINS AND ENZYME KINETICS

The proteins comprise one of the most important classes of macromolecules found in living organisms. The basic monomers which polymerize into a protein structure are the amino acids with formula $H_2N—CHR—COOH$, the R being one of twenty commonly found groups. The polymer structure has the form of a linear chain, the so-called polypeptide chain or a combination of several of such chains linked together with disulfide or hydrogen bonds. The molecular weight of protein molecules varies from 6500 for trypsin inhibitor to 11,466 for insulin to 40×10^6 for tobacco mosaic virus (although the TMV structure has 2130 subunits, which means only $\sim 2 \times 10^4$ per subunit). Myoglobin has a molecular weight of 17,000, which is about the order of each subunit of a protein composed of many subunits. Thrombin and fibrinogen, which will be discussed



later in connection with blood coagulation, have molecular weights of 31,000 and 340,000, respectively.

The rate at which the elucidation of protein structures has been developing can be appreciated by noting that the first complete protein composition analysis was made by Erwin Brand et al.,[1,2] in 1945. They determined the number of each of the various amino acid residues in β-lactoglobulin. The number of proteins so analyzed is now in the hundreds.[3,4] The breakthrough came with the introduction of ion exchange chromotography in the late 1950s, after which the determination of the amino acid content of proteins became routine.[5,6]

The first sequence analysis in which the order of amino acid residues was given was completed in 1951 by F. Sanger and collaborators,[7] who obtained the ordering for bovine insulin. By now each year brings forth new lists of complete sequences. These are generally compiled in the *Atlas of Protein Sequences*.[8] The total number of proteins sequenced about doubles each year. The sequences that occur in some proteins found in different mutants of the same organism have even been compared.

The first complete three-dimensional X-ray structure analysis of proteins was made by Kendrew[9] and Perutz[10] for myoglobin and hemoglobin in the early 1960s. Although such structure determinations are still difficult, many more have been completed. A discussion of these analyses forms the complete proceedings of the 1971 Cold Spring Harbor Symposium.[11] One of the most complicated structure determinations made to date is that of the complex of the antibiotic actinomycitin with DNA.[12]

Of course many general structural aspects of proteins had been deduced before the first complete X-ray structure was determined. Crudely speaking, it was observed that there are two classes of proteins, the fibrous ones which tend to form the structural materials of organisms, and the globular ones which tend to be more functional in participating in chemical reactions. Since some of the fibrous proteins such as those of membranes and muscle fibers are both structural and functional, and not all the globular proteins are completely spherical, the class division is not very precise. Portions of a globular protein generally have a helical form (the α-helix conformation[13]). The fibrous proteins generally contain sections of sheets (the β-sheet conformation) of protein chains held together by hydrogen bonds. The chains are generally bonded together in an antiparallel manner (with the directed axes alternating between successive chains). Occasional exceptions exist with parallel binding. Two good reviews of biopolymers have recently been published, one a review of structural characteristics by B. Jirgensons[14] and the other a book by F. Wold on both structure and function.[15]

The nomenclature for various levels of protein structure is as follows.

Primary structure is used to describe the covalently linked amino acid residues; *secondary* structure refers to the weak interactions between the nearest-neighbor groups and the resulting local geometrical structure; the *tertiary* structure is the total folding of the molecule into a unit; and the *quaternary* structure is a statement of how various subunits are connected. Proteins sometimes involve heavy metals and other unusual and specialized groupings which are called nonprotein prosthetic groups (for example, iron, in hemoglobin).

The special proteins that are the subject of this review, the enzymes, are catalysts of chemical reactions in living systems. Since the control of biological functions is almost entirely through chemical processes, the enzymes are central components of the control system of an organism. The concept of an enzyme as a catalyst goes back to the turn of the century. The classical schematic form for an enzyme reaction is (S representing substrate which, in the presence of an enzyme is converted to a product P)

$$S \xrightarrow{E} P \tag{1}$$

Further details of the process were introduced in the Michaeles-Menten model,[16] which is characterized by the equation

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C \xrightarrow{k_2} E + P \tag{2}$$

with C being an unstable complex which either reverts back to the substrate form or continues to develop into a product, freeing the enzyme. This complex is a molecular entity not to be confused with the Eyring activated complex which is often included in the elaboration of the character of enzyme reactions.[16]

The classical theory of enzyme kinetics is reviewed in a number of places.[15–17] The basic postulates used in enzyme kinetics are as follows:

1. Enzymes are perfect catalysts that are neither consumed nor transformed in their catalytic activity.

2. Enzymes possess varying degrees of specificity toward compounds with which they react; some enzymes act upon large groups of similar compounds; some appear to act upon only one compound, but the specificity is never perfect. Hence analogues of natural substrates may mimic them, by combining with enzymes to form stable complexes which immobilize the enzymes and prevent them from performing their normal function.

3. All enzymes known in nature are proteins. The enzyme function depends on its conformation so that any action which changes its geometrical shape may also change its catalytic activity.

These characteristics suggest a number of properties of enzymes (actually, the experimental observation of these properties suggested the above postulates). Although the rate of most chemical reactions increases with temperature, enzymatically catalyzed reactions do this only to a certain critical temperature beyond which the rate decreases. One of the reasons for the rate increase is that the velocity of reactants increases with temperature, thus increasing collision frequencies. However, enzyme macromolecules assume more random configurations and suffer the breaking of hydrogen bonds with increasing temperature. At sufficiently high temperatures, the tertiary structure required for catalytic behavior becomes so distorted that the reaction no longer can occur.

The existence of tightly bound inhibitors to enzymes corroborates the traditional concept that the catalytic character of the molecule is localized at an "active center." Indeed, by separating the part of the enzyme attached to the inhibitor from the rest of the chain and then separating the remaining enzyme fragment from the inhibitor, the nature of the localized center could be determined[18] chemically. If a heavy metal marker could be attached to the inhibitor one might use X-ray or electron microscopic techniques to localize the center. Recently it has been noted that the geometry of water molecules attached to active sites can be investigated by proton magnetic relaxation methods.[19–23]

Since enzymes act as catalytic molecules in living systems, it is clear that they play an important role in biological control mechanisms. By changing the tertiary structure or by adding or withdrawing an inhibitor, one can monitor the rate of production or destruction of molecules with control functions. Now it is well known that the degree and precision of control of any process depends on the number of elements which interact with each other in the control mechanism, on various time lags which are natural in the system, and on various feedback loops which appear in it (a review of various biological control mechanisms with a good bibliography is given in Ref. 24). Hence one would expect to find examples of enzymatic processes in which a number of enzymes participate, each controlling one component of a large control network.

One of the most delicate control mechanisms required in vertebrates is the control of blood coagulation. The uninterrupted, well regulated control of blood flow is crucial for life in that the bloodstream transports food and oxygen, manages waste collection, and controls temperature and water flow. Since an organism must be able to survive minor wounds, the circulatory system must through clotting be able to prevent extensive blood loss. At every moment, at every point in the system, the circulatory system has to make the decision; should clotting start or should it not. Clotting during normal flow in the absence of wounds is as serious a

fault as not clotting when necessary at the site of a wound. It has long been known that many "factors" participate in the clotting mechanism. The function of the many factors is probably to make a fail-safe system which responds only at the proper time and to provide an amplifying cascade process for the generation of the appropriate chemical reactions to complete the clotting process.

In the process of trying to organize the relation between and the action of the various clotting factors, R. G. MacFarlane[25] introduced the concept of an open enzyme cascade or triggered enzyme system as defined in the diagram

$$\begin{array}{ccccc} & & E_1 & & \\ S_1 & \longrightarrow & E_2 & & \\ & & \downarrow & & \\ & S_2 & \longrightarrow & E_3 & \\ & & & \downarrow & \\ & & S_3 & \longrightarrow & E_4, \text{ etc.} \end{array} \qquad (3)$$

in which the jth reaction is catalyzed by the jth enzyme to produce the $(j + 1)$th enzyme. It is clear that such a reaction network could have tremendous amplifying properties. A small amount of E_1 induces the production of a larger amount of E_2, which induces the production of a still larger amount of E_3, etc.

A cascade that might be more sensitively controlled is the damped cascade,[26,27] in which each enzyme can be inactivated by a change in form

$$\begin{array}{ccccccc} & & E_1 & & & & \\ S_1 & \longrightarrow & E_2 & \longrightarrow & I_2 & & \\ & & \downarrow & & & & \\ & S_2 & \longrightarrow & E_3 & \longrightarrow & I_3 & \\ & & & \downarrow & & & \\ & & S_3 & \longrightarrow & E_4 & \longrightarrow & I_4, \text{ etc.} \end{array} \qquad (4)$$

The remainder of this review is devoted to describing a number of enzyme cascades and briefly outlining the theory of control aspects of such cascades.

II. DESCRIPTION OF SEVERAL ENZYME CASCADES

A. Blood Coagulation

Since the importance of enzyme cascades was first emphasized in connection with blood coagulation, we start a set of descriptions with that example. Two independently conceived papers, the one mentioned above by MacFarlane[25] and the other by Davie and Ratnoff,[28] described clotting as a cascade process. The titles of the respective papers are "An Enzyme Cascade in the Blood Clotting Mechanism and Its Function as a Biochemical Amplifier" and "Waterfall Sequence for Intrinsic Blood Clotting."

The original figure from each paper is reproduced here. The history of the pattern of observations which led to the cascade idea is outlined by MacFarlane[29] and by Laki in his book on fibrinogen.[30]

There are two main features of the cascade. The first, which involves the larger number of "factors," is the preparation or buildup stage (which, for reasons discussed below, will be called the decision-making stage); and the second, which starts with the production of thrombin and ends with the formation of the hemostatic plug that stops the leakage of the blood, is a polymerization and stabilization stage.

The classical theory of clotting, developed by Morawitz in 1905, involved four factors; prothrombin which, under conditions appropriate for clotting, reacted with the thromboplastin and calcium to produce thrombin which catalyzed the polymerization of fibrinogen molecules (normally present in the blood) into fibrin, the structural component of the clot. Of course in 1905 there was little understanding of the size, shape, or structure of the relevant molecules.

A test that became standard for assaying blood was developed by Quick[31] in 1935. It essentially related the prothrombin concentration to the time required for clotting. Owren,[32] in 1947, in applying this test to a patient's blood with an excessively long clotting time, noticed upon chemical analysis of the blood that the prothrombin concentration was not abnormal. Since the clotting time of the patient's blood was reduced to a normal value when normal plasma from another individual was added to the blood, Owren concluded that his patient lacked a hitherto unexpected clotting factor which he called factor V.

Once the new factor was proposed, it was natural that the blood of hemophiliacs should be characterized by the absence of another clotting factor, which was called VIII. It was not long before a number of patients with abnormal clotting characteristics appeared at various physicians' offices. The abnormalities of patients Hageman, Christmas, and Stuart were considered to be due to the absence of three other factors which were named after these individuals. By 1964 about twelve clotting factors were deemed to be necessary and certain relations seemed to exist between them. For example, it was noted by Kingdon, Davie, and Ratnoff[33] that factor XII exhibited enzyme activity such that when factor XI reacted with active XII, an enzyme was produced which caused the activation of factor IX. Biggs, MacFarlane, Denson, and Ash[34] then noted that after reacting with factors XI and XII, factors VIII and IX somehow produced an enzyme which actuated factor X. Davie and Ratnoff noted that of particular importance is the concept that the protein clotting factors interact in pairs in which one member acts as an enzyme and the other as a substrate. As a result of these and other similar observations,

Davie and Ratnoff[28] and MacFarlane[25] independently proposed that clotting is a result of the sequence of events listed in Figs. 1 and 2.

After the publication of the clotting cascade scheme, several groups investigated the details of the chemistry of the scheme as well as the consistency with clinical experience. The results of the biochemical research have been summarized by M. P. Esnouf.[35] The procedure used was to attempt to isolate the various relevant factors from blood to see if the individual reactions in the cascade occurred as proposed. Since each step in the cascade is an amplifier, those substrates that appear at late stages of the cascade process exist in higher concentrations in the blood plasma and can be separated out easily. Fibrinogen and prothrombin are thus available for experiments. Thrombin can be prepared from prothrombin by the action of proteolytic enzymes such as trypsin, or by physiological activators and recovered from the reactants by chromatography.[35,36] Kezdy, Lorand, and Miller[37] have determined that there is one active center per 32,600 g of thrombin. Factors Xa and V have also been separated.[38,39] Bovine factor XII has been isolated by Schoenmakers et al.;[40] however, experiments directly verifying that this factor plays the role assigned to it in the cascade are not yet available. Esnouf[35] concludes that from "results obtained from biochemical investigations of the clotting mechanism, it can be seen that evidence is emerging that some factors are zyomogens and are transformed into active enzymes, which is consistent with the 'cascade hypothesis,' while other factors such as factor V and possibly factors VIII and VII act as accelerators for specific reactions." Seegers, who has proposed an alternative model for clotting which will be discussed below, is doubtful as to whether all the enzymes listed in the cascade exist.[41]

Rosemary Biggs[42] has critically reviewed the clinical evidence for the cascade model. Such evidence is important because without clinical experience with patients with clotting abnormalities, there would have been no motivation for the development of the cascade model. Although the original interpretation of abnormally long clotting times was associated with the absence of some factor, it might also have been interpreted as the existence of abnormal inhibitor activity, since many enzyme reactions are known to be influenced by naturally occurring inhibitors in vivo. Indeed heparin, a liver extract, was noted by Howell and McLean in 1916 to be an anticoagulant, its specific action being the inactivation of thrombin. The detailed process might be an interplay between factors and inhibitors, which could be characterized by the damped cascade of (4).

Clotting time tests, which are commonly used clinically, provide a means for examining the role of various factors. Different steps in the cascade involve different time scales, the first reaction taking the longest

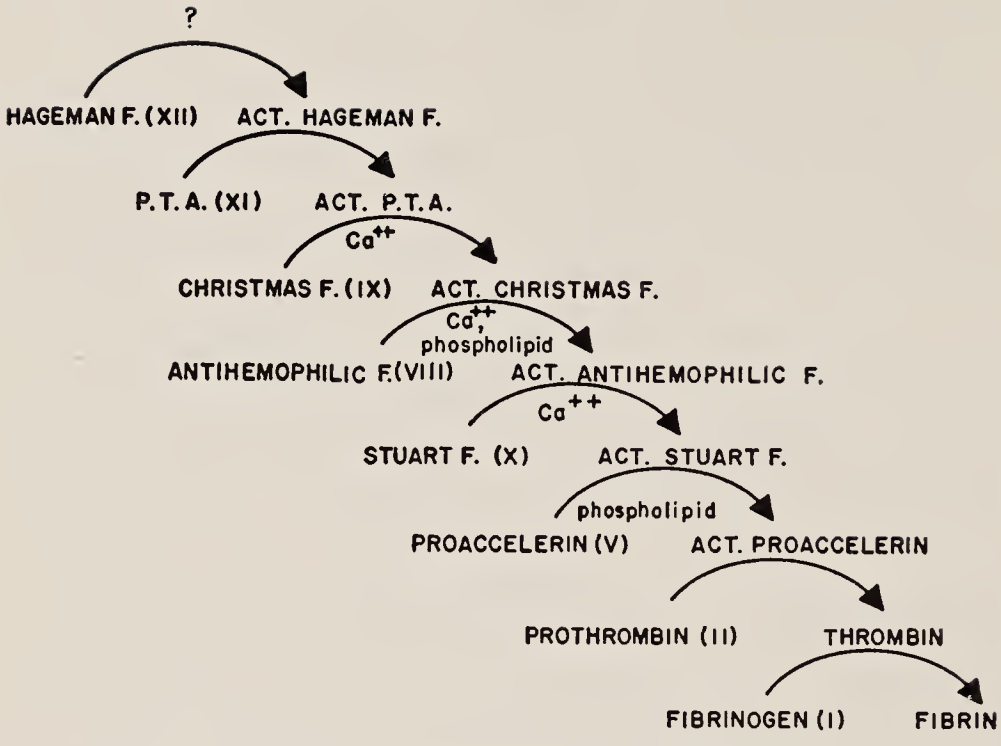


Fig. 1. The Davie and Ratnoff "waterfall" sequence for intrinsic blood clotting as diagramed in *Science*.[28]

time and the late stage in the cascade, the polymerization of fibrinogen to fibrin, taking the shortest time. One way of estimating[42] the time required for each step is to add appropriate amounts of material that is known to cause clotting by bypassing various factors and to see how the clotting time is reduced by the bypass.[42] Biggs noted that when kaolin is added to plasma, factors XI and XII are rapidly activated. Brain extract bypasses

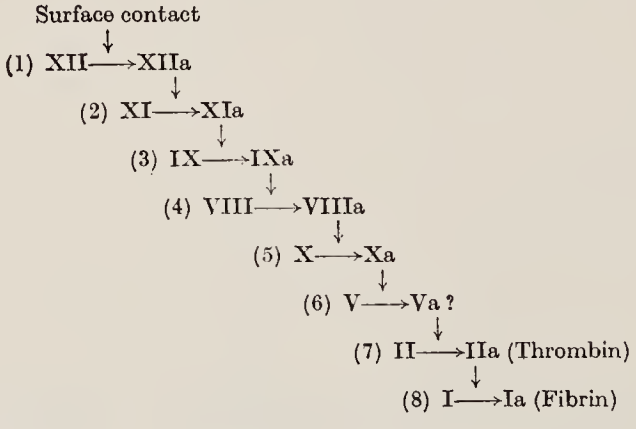


Fig. 2a. MacFarlane's enzyme cascade for clotting initiated by surface contact diagramed in *Nature*.[25]

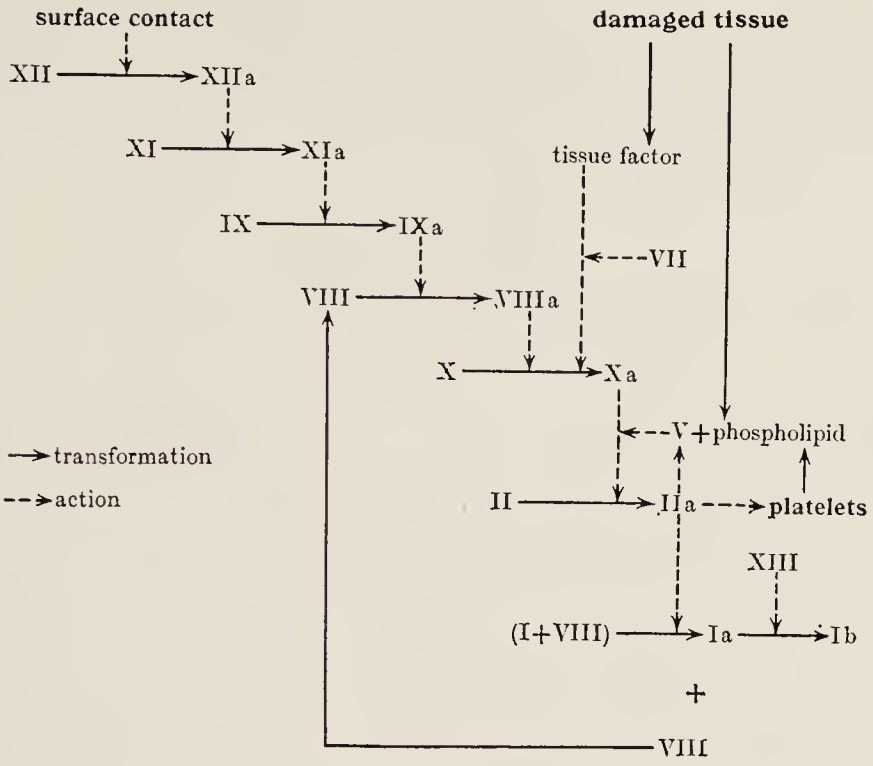


Fig. 2b. MacFarlane's clotting cascade as it develops from either surface contact or tissue damage. Reproduced from *Proceedings of the Royal Society (London)*.[29]

factors VIII, IX, XI, and XII. Addition of Russell's viper venom causes the bypassing of all factors preceding X. Finally the addition of concentrated thrombin to plasma starts the fibrinogen polymerization almost immediately. Table I (from Ref. 42) shows the reduction in clotting time of normal plasma upon various bypassing additions. The facts support the cascade hypothesis.

Two-stage tests have also been organized in which the cascade is divided into two parts such that in one part the coagulant is formed and in the

TABLE I
One-Stage Clotting Time for Normal Plasma with Bypassing Additives

Additive	Clotting Time in Si Tube (sec)
None	250
Kaolin	45–70
Brain extract	12–15
Russell's viper venom	5–7
Much thrombin	2–3

other part it is tested.[42] Results of such tests seem also to be consistent with the cascade hypothesis. Finally, tests have been made with the introduction of inhibitors of specific enzyme reactions to see if the cascade is interrupted at the appropriate points. The results of these experiments seem to be in accord with the cascade theory.

Clinical practice can be related to experiments by assessing the correlation between diagnostic tests and severity of blood disorder and by relating laboratory tests to control of patient treatment.

Diagnostic tests can indicate the abnormally low level of various factors. The clinical consequence of these low levels has been observed as follows.[42] Strangely, almost complete absence of fibrinogen is not lethal although with no fibrinogen there are no clots. Complete absence of prothrombin has never been encountered but a level as low as 10% of normal is exceedingly dangerous and even 40% of normal amount is still a problem. Thus thrombin deficiency is much more severe than that of fibrinogen. Very low levels of factors V, VII, and X lead to bleeding disorders of variable severity but a reasonable life is still possible at even as low as 1% of normal abundance. Deficiencies of VIII and IX cause defects similar to classical hemophilia; that of XI is accompanied by mild and unpredictable bleeding, while a lack of XII is not associated with abnormal bleeding. Biggs concludes that, "if blood clotting is essential to hemostasis, then the sequential hypothesis does not fit well with clinical observations. The facts suggest that if blood fails to start clotting or fails to finish the process off the defect is less noticeable than if the chain is interrupted in the middle." On the other hand, she notes that tests for factors I, II, V, VII, VIII, IX, and X are well correlated with treatment. Laboratory tests for XI and XII are useful for diagnosis, but the results are not related to severity of bleeding.

There is a school of thought which believes that platelets and not blood clotting form the basis for hemostasis. Hayem and Bizzozero almost a hundred years ago showed that platelets have a unique property of sticking to the edges of cut blood vessels. They also cohere to each other,[43] forming clumps that are certainly an essential part of the hemostatic plug. Recent work[44,45] suggests that adenosine diphosphate, ADP, causes platelets to adhere to each other; ADP is released from platelets by thrombin so that under the appropriate conditions on appropriate surfaces there is platelet adherence and a form of platelet fusion known as viscous metamorphosis occurs.

Platelet action also seems to be a cascade process of sorts with twelve platelet factors having been identified.[46] The scheme of platelet role in hemostasis as presented by Shirley Johnson in Ref. 46, which is an excellent review of the subject, is given in Fig. 3.

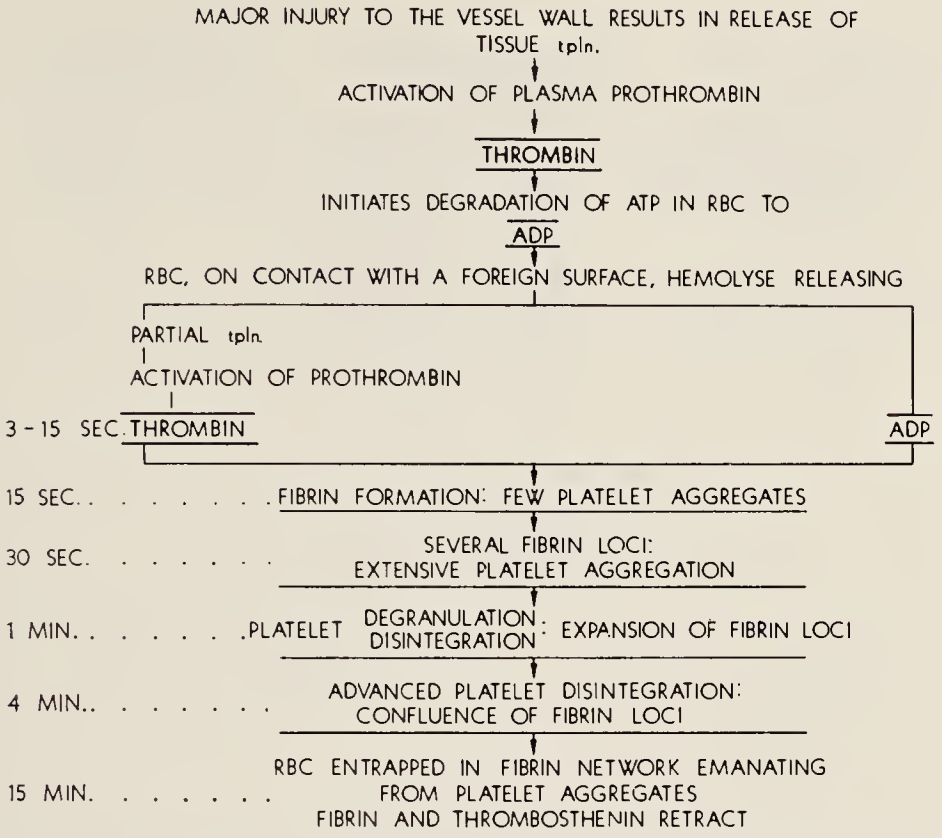

MAJOR INJURY TO THE VESSEL WALL RESULTS IN RELEASE OF TISSUE tpln.
ACTIVATION OF PLASMA PROTHROMBIN
THROMBIN
INITIATES DEGRADATION OF ATP IN RBC TO
ADP
RBC, ON CONTACT WITH A FOREIGN SURFACE, HEMOLYSE RELEASING
PARTIAL tpln.
ACTIVATION OF PROTHROMBIN
3 - 15 SEC. THROMBIN
ADP
15 SEC. . . . . FIBRIN FORMATION: FEW PLATELET AGGREGATES
30 SEC. . . . . SEVERAL FIBRIN LOCI: EXTENSIVE PLATELET AGGREGATION
1 MIN. . . . . PLATELET DEGRANULATION: DISINTEGRATION: EXPANSION OF FIBRIN LOCI
4 MIN.. . . . . ADVANCED PLATELET DISINTEGRATION: CONFLUENCE OF FIBRIN LOCI
15 MIN. . . . . RBC ENTRAPPED IN FIBRIN NETWORK EMANATING FROM PLATELET AGGREGATES
FIBRIN AND THROMBOSTHENIN RETRACT

Fig. 3. The role of platelets in hemostasis as outlined by Shirley Johnson. Reproduced by permission of Academic Press.[46]

Although platelets are clearly important it also seems clear that they alone are insufficient for hemostasis since many patients with normal platelets but defective coagulation may have defective hemostasis.[42] Mustard et al.[47] have found that in hemophiliac dogs the platelet masses form at wounded vessel sites but that they are not firmly compacted and are easily washed away. Hence it seems that the fibrin network is necessary to stabilize the platelet masses.

Seegers[41] has emphasized a point of view of coagulation rather different from that of the cascade school. He elaborates on the classical model of Morawitz in which the basic process is still the transition of prothrombin to thrombin. The crucial feature in his mechanism is the delicate balance between a large number of procoagulants and anticoagulants which tend to generate thrombin and retard its generation, respectively. A list of a number of compounds in each class is given in Fig. 4, which summarizes this mechanism. Seegers[41] has interpreted various plasma deficiencies and hemophilia in terms of abnormal prothrombin molecules.

The molecular details of the last phase in the clotting process, the

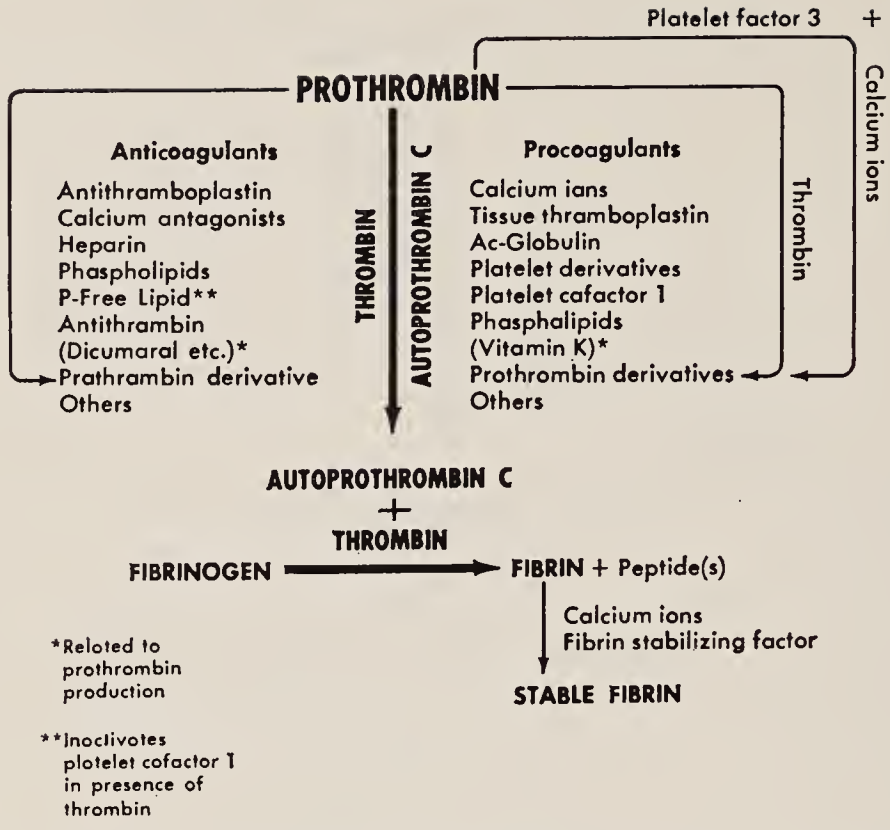


Fig. 4. Seegers' elaboration of the classical four-factor model of clotting. In this picture there is a delicate balance between coagulants and anticoagulants. When the coagulants win out, as they should under normal clotting conditions, thrombin is formed and the clotting process begins. Reproduced by permission of Academic Press.[41]

polymerization of fibrinogen to fibrin, is more clearly understood than that of the preliminary decision-making process. Although this is not our primary interest here, it is the conclusion for which the decision making evolved and therefore deserves some attention.

Early electron micrographs of fibrinogen[48,49] indicated that it was composed of three protein spheres or nodules, each about 50 to 60 Å in diameter, connected by a thin strand 8 to 15 Å thick, with the full length of the molecule being about 475 Å. G. Koppel,[50,51] from more detailed and more modern electron micrographs, has concluded that the nodules are pentagon dodecahedrons. Still further structural analysis shows the molecule to be a dimer with the individual parts of each monomer being linked by bisulfide bonds, as indicated[53] in Fig. 5.

A number of investigators including L. Lorand[52] and A. G. Loewy[54] and their colleagues have, over the past twenty years or more, developed a

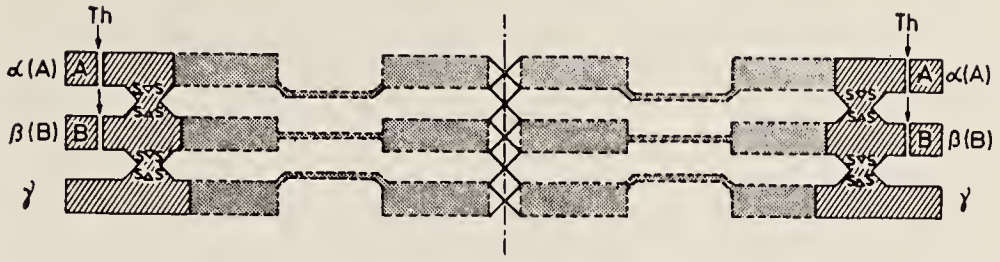


Fig. 5. A schematic sketch of the structure of the fibrinogen dimer. Polymerization into fibrin is effected by the cleavage of the (A) and (B) polypeptide ends by thrombin. Reproduced from *Thrombosis et Diathesis Haemorrhagica*.[53]

consistent picture of the major steps in the polymerization process which ends with a crosslinked stabilized clot. The steps have been summarized by Lorand in Fig. 6 and the important events are sketched geometrically in Fig. 7. The ternodular rod-shaped fibrinogen dimer is normally negatively charged so that the molecules repel each other as they circulate in the bloodstream. The first step in the polymerization process is the anchoring of the thrombin to the polyionic charges near the ends of the molecule. The α(A) and β(B) polypeptide ends of the molecule are cleaved by the thrombin as indicated in Fig. 5. The cleavage occurs at amino acid positions 16 and 19.[53] Upon reduction of the total charge by the removal of the charges at the ends, the fibrinogen molecules clump together into a clot. The sites that will be available on the protein for reaction with the ligase enzyme and crosslinking are then exposed. The enzyme is called factor

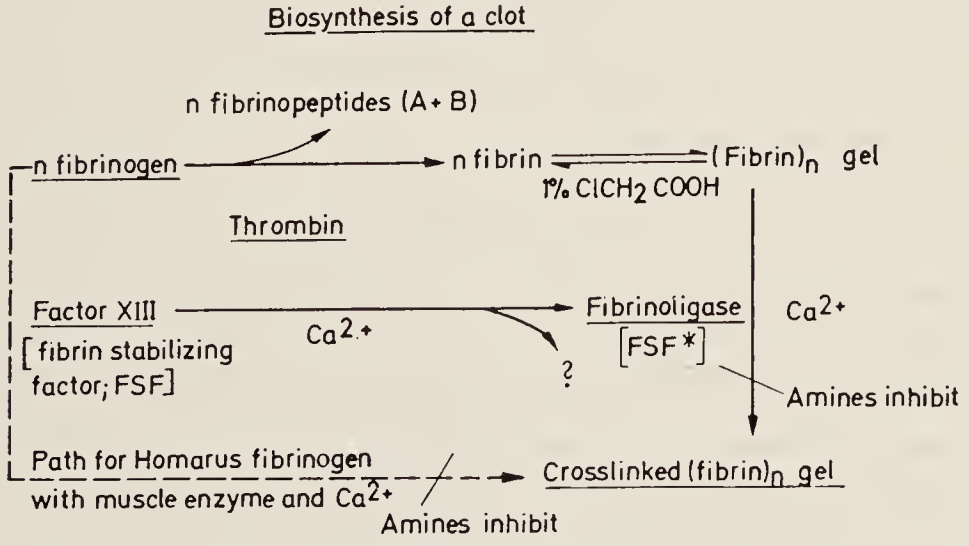


Fig. 6. Steps in the polymerization of fibrinogen to fibrin and the formation of the final crosslinked fibrin gel, according to Lorand.[52] Reproduced from *Thrombosis et Diathesis Haemorrhagica*.

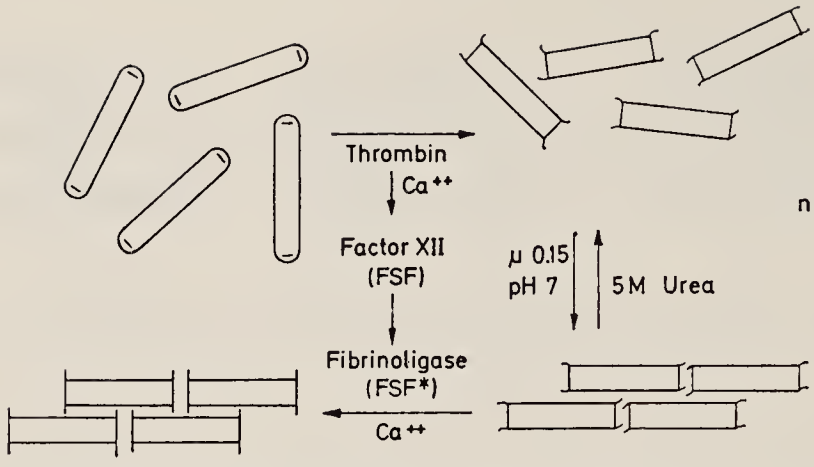


Fig. 7. Configurational changes associated with the steps listed in Fig. 6. Reproduced from *Thrombosis et Diathesis Haemorrhagica*.[52]

XIII, the fibrin stabilizing factor. The crosslinking is completed through γ-glutamyl-ε-lysine bonds. The detailed molecular structural changes and the molecular character of inhibitor reactions have been reviewed.[52] Many beautiful electron micrographs of the fibrin clot have been taken by a number of workers in the field. Some of these have been collected by Bang[55] in his article on the ultrastructure of the fibrin clot.

Ferry and his associates[56,57] have presented models of the gel formation and a staggered overlapping rod structure for the final polymer. Ferry has also discussed the mechanical properties of both ligated and unligated clots. The shear modulus of unligated clot, measured in a time scale of the order of 10^{-2} sec, is roughly proportional to the square of the volume fraction of fibrinogen with a proportionality constant of the order of 10^6 dyne/cm^2. The relaxation modulus of ligated clots is nearly constant over a very long time interval while unligated clots show stress relaxation within minutes.

The amino acid residue sequence analysis of fibrinogen has been motivated by interests in the mechanism of evolution as well as by a desire to understand the clotting process. Blombäck[58] states that fibrinogen is a suitable protein for taxonomic studies in closely related mammalian species because its primary structure has changed relatively quickly during evolution. Thus it has been estimated[58] that the rate of mutation per 100 amino acid residues in certain segments of fibrinogen is 90 per million years, whereas the corresponding figures for some other much studied proteins, the hemoglobins and cytochrome C, are 12 and 13, respectively.

As Eck and Dayhoff organized data in the literature on protein sequences[8] they noticed tremendous similarity in the sequences in homologous proteins from diverse species. For example, in the case of cytochrome C, a

protein of the order of 100 to 110 amino acid residues (the exact number depending on the species), 35 appear at points common to all 19 species examined. There were 33 locations at which one of two possible residues were found, etc. The position in the chain with greatest variation was 96, at which eight different residues were observed in the different species. The longest sequence common to all species is that which extends from position 76 to position 86. One might be tempted to conjecture that that sequence contains the functional unit of the molecule.

An examination of the sequences of a given protein which correspond to various species shows that some species have exactly the same sequence. From this it is concluded that they had a "common ancestor" with that sequence. A number of now extinct common ancestors can then be listed. Although the sets of "common ancestor" differ from each other in amino acid arrangement, they can be ordered so that the most similar ones are placed near each other. Through this ordering an evolutionary tree of organisms, real and extinct, can be constructed so that those that have similar sequence are close to each other on the tree. These qualitative ideas have been made more precise in the work of Eck and Dayhoff and evolutionary trees have been constructed.

B. Blombäck and associates[53] and Doolittle and his associates[59,60] have determined long amino acid residue sequences for fibrinopeptide A and fibrinopeptide B, two important regions in fibrinogen. They have located the sites upon which thrombin acts to prepare the molecule for the polymerization process. The evolutionary tree exhibited in Fig. 8 has been derived[8] from these sequences. Work is now in progress on the characteristics of abnormal inherited human fibrinogen which does not coagulate properly.[61,62] Such examples are referred to in the literature as fibrinogen Baltimore, fibrinogen Zürich, fibrinogen Detroit, etc. The structural defect in fibrinogen Detroit seems to be well understood.[63]

Since sequence analysis has now begun on thrombin and prothrombin,[64] one can expect that a deeper understanding of their roles in stimulating the polymerization of fibrinogen will develop.

Most research on hemostasis has been done to understand the natural process and to develop clinical procedures for treating abnormalities. With the present concern with organ transplantation, and replacement of heart valves and segments of arteries and veins, there is developing an interest in the influence of foreign substances on clotting. One of the main bottlenecks in the replacement program is that neither nonthrombogenic nor hypothrombogenic surfaces are known, so that clots tend to form spontaneously when there are foreign surfaces in the circulatory system. Electron micrographs have been taken of blood components adherent to surfaces of hevea rubber, Teflon, and silastic rubber.[65] The effects of the

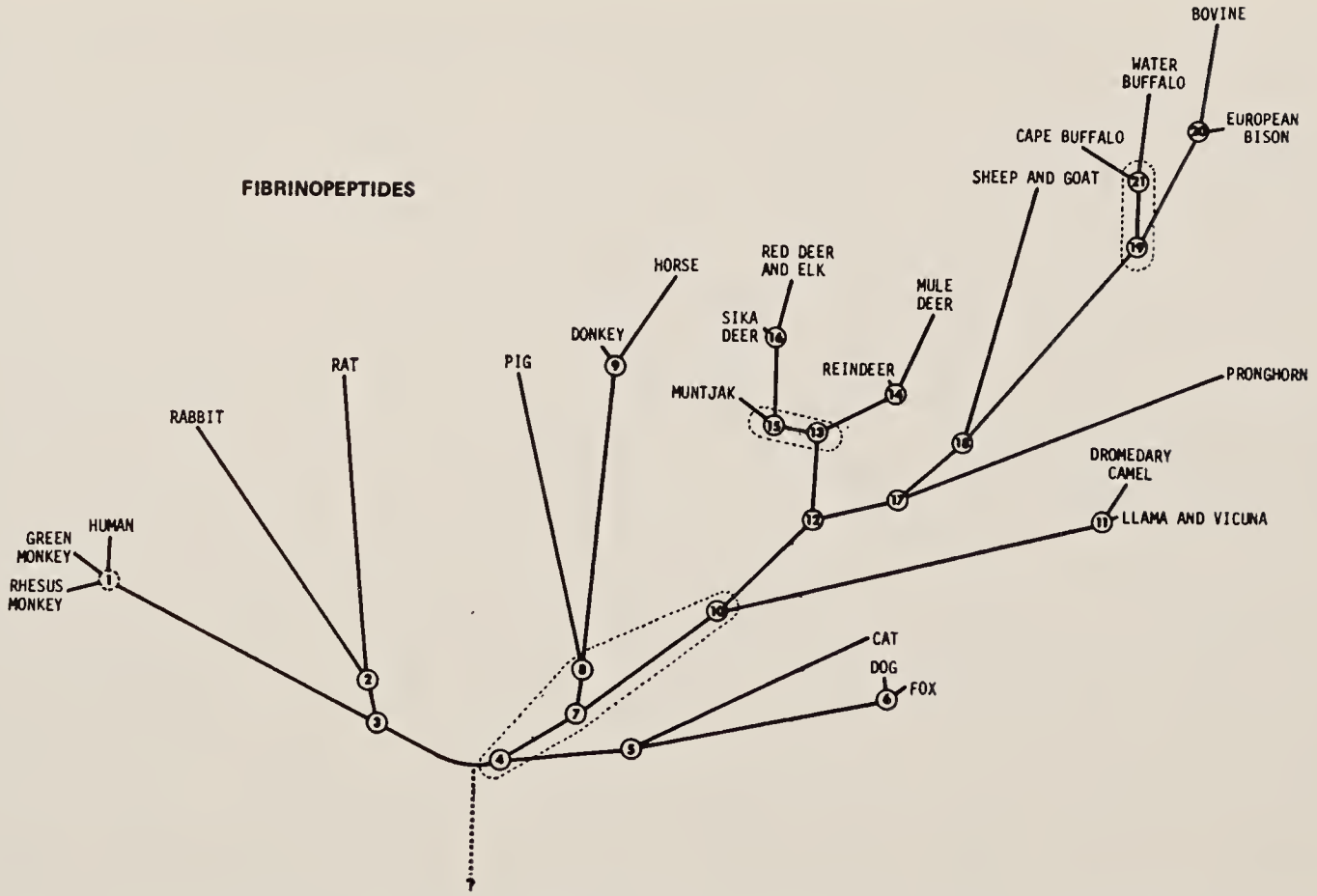


Fig. 8. An evolutionary tree derived from comparing amino acid sequences in active parts of fibrinogen molecules from various species[8,28]. Figure[28] reproduced by permission of North Holland publishers.

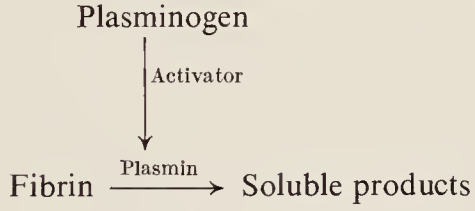
chemical structure and surface properties of synthetic polymers on blood coagulation are discussed in Ref. 64; platelet adhesion induced by fibrinogen absorbed on glass has been investigated,[66] as have the initial events in the interaction of blood with foreign surfaces.[67,68]

The review of several other enzyme cascades in blood plasma, which follows, is somewhat shorter than that of clotting.

B. Fibrinolytic Mechanism

Very small clots are frequently formed to seal tiny deficiencies in the endothelium and are occasionally formed spontaneously in an accidental manner in the bloodstream. In a well designed control system, a response should be available to digest these when they are not wanted. It has been suggested[69] that the plasminogen-plasmin system for fibrinolysis may be coupled to the clotting system to play this role for the maintenance of an intact vascular tree in an organism. According to this hypothesis the coagulation system seals deficiencies in the endothelium while the fibrinolytic system removes such deposits after they have outlived their usefulness.

Fibrinolysis is the digestion of fibrin by an enzyme which converts it into components that are soluble in the blood plasma. The plasminogen-plasmin system which performs this function operates according to the diagram below.



Plasminogen, a plasma β-globulin, has a molecular weight of about 89,000; the amino acid composition of human plasminogen has been reported by Robbins et al.[70] There are 22 bisulfide bonds in the plasminogen molecule. Wallén[71] has determined short amino acid sequences at the N-terminal end of the molecule. In the presence of an appropriate activator it is converted to plasmin, a proteolytic enzyme. It is an endopeptidase which digests fibrinogen, fibrin, factor V, factor VIII, prothrombin, ACTH, growth hormones, some members of the complement system, etc.[69,72] Its conversion of fibrin to soluble products is similar to the action of thrombin and trypsin in that they all attack the *l*-argynl and *l*-lysyl bonds inside peptide chains.[73]

The understanding of the character of the activation and indeed of the process itself started with some interesting observations made over many

years. For example Hunter in 1794 noticed that the blood of animals killed by lightning or after being exhausted in the hunt did not coagulate. There are more recent relevant observations that in man fibrinolytic activity can be provoked by surgical operations,[74] and that certain strains of hemolytic streptococci contain a substance capable of lysing human blood clots.[75] In 1889, Denys and Marbaix[76] found that the treatment of serum with chloroform induced fibrinolytic activity. All these observations motivated detailed biochemical research programs which have led to the conclusion[69] that the activation of plasminogen can be accomplished in four main ways: (*1*) by substances such as urokinase (a molecule present in normal urine) which can activate plasminogen directly; (*2*) by substances such as streptokinase and tissue lysokinase which convert a normally inert proactivator into activator; (*3*) by substances such as chloroform which activate in vitro; and (*4*) by administration of nonactivators which stimulate the release of activators in vivo, for example, nicotinic acid. Certain tissues are rich in activator, especially those concentrated around blood vessels.

Normal plasma seems to contain inhibitors to the activation of the plasminogen-plasmin system,[77] and there are a number of chemical agents that have been discovered to have the same property. Some of these are the amino acids such as ε-aminocaproic acid (EACA), aminomethylcyclohexane-4-carboxylic acid (AMCHA), and *p*-aminomethylbenzoic acid (PAMBA). A polypeptide extract of beef parotid also inhibits plasmin activity. Inhibitors have been administered to patients who undergo serious postsurgical fibrinolysis.

A clinical observation favoring the hypothesis of the linking of fibrinolysis and clotting is that degraded fragments of fibrin and fibrinogen can be detected in patients who suffered from episodes of intravascular fibrin formation.[78] In some instances[79] the amount of degradation product indicates that over half of the total circulating fibrinogen has been lysed or converted into fibrin and then lysed. On the other hand,[69] it has not been clearly established that fibrinolytic breakdown products exist in normal serum.

C. Complement and the Complement Cascade

Antibodies have the function of recognizing the foreignness of substances (antigens) which have invaded the body. Generally the antibody attaches itself to the antigen to form a complex. "Complement" is the series of reactions which is activated by the formation of the complex and which results in the destruction of the invader.

The complement reactions seem to perform three functions;[80] they (*1*) kill cells directly, (*2*) enhance phagocytosis of cells, and (*3*) liberate vasoactive substances. The direct killing of invading cells is done by the products of the complement reactions weakening the cell walls of the invader[81,84] so that osmotic control of the cell is lost and the cell ruptures. It has been observed that in certain cases neutrophil leucocytes will engulf foreign red cells coated with antibody and the products of complement but not merely the red cell antibody complex.[80] Hence in some cases complement enhances phagocytosis. Finally, the activation of complement may produce "anaphylatoxin," which in turn liberates vasoactive substances from various cells.[80,85] It is suspected that this causes local slowing of the circulation to help localize antigen-antibody complexes.

The complement system involves a number of proteins, many of which are fairly well characterized. As with blood coagulation, the complement system is essential when invasion of the body occurs but can lead to ill effects if started when there is no invasion, or if it is active throughout large regions when the invasion is local. There are of the order of nine proteins involved, some of which have distinct subunits. These proteins are identified by the names C′1, C′2, . . . , C′9 with the subunits of C′1 being C′1q, C′1r, and C′1s, etc. The complement cascade is probably not linear but has some branching steps with feedback. The details of the cascade are presented in Refs. 82 and 83. We now give a very brief summary of them.

Attachment of the antibody to the antigen initiates the fixation of the first complement protein C′1 to the antibody, subunit C′1q carrying the site of fixation. The three subunits of C′1 are held together in a complex in the presence of calcium ions. The natural substrates catalyzed by C′1 are not in the cell wall, but are the two succeeding complement components C′4 and C′2. When C′4 is acted upon by C′1 esterase it attains a short-lived activated state which can bind to the antigen membrane, the antibody, or other C′4 molecules. Activated C′2 binds to C′4 in the presence of Mg ions. A substantial fragment of this bound enzyme is cleared by activated C′1 with C′42a or C′3 convertase resulting. To this stage no biological effects are produced in the complement cascade.

The fixation of the C′3 which occurs at this point is analogous to the polymerization of fibrinogen into the fibrin form. As many as 1000 molecules of C′3 can be fixed for each bound C′4. A rapid succession of reactions which involve C′5, C′6, and C′7 then occurs with a complex forming that retains activity for many hours, even at body temperature of 37°C. This complex seems to yield the priming lesion in the antigen. The final weakening of the cell membrane which in the end kills the cell is accomplished with components C′8 and C′9.

D. The Renin-Angiotensin System

One of the characteristics of the circulatory system which must be controlled is the blood pressure. After exercise, during which the blood pressure increases considerably, the abdominal muscles relax and blood is drained into the venous reservoir. However, when the pressure gets abnormally low the response seems to be a more subtle one that involves an enzyme cascade.

When there is a reduction in blood flow through the kidney an enzyme called renin (mol wt 40,000) is released from the kidney. This in turn acts upon a substrate in the blood plasma to produce a pressor substance called angiotensin.[86] The sequence of events seems to be that the renin released into circulation acts upon the substrate to produce a biologically inactive decapeptide which is converted[85,86] into an active octapeptide by cleaving of the two C-terminal amino acids His-Leu. The components and interactions of the renin-angiotensin conversion are shown in Fig. 9.

Physiologically it seems that the angiotensin acts directly on smooth arterial muscle causing it to contract, thus raising the blood pressure. A second effect is that it causes a reduction in sodium excretion. It has also been shown[86] that angiotensin stimulates the secretion of aldosterone.

Renin has a long half-life in the circulation (> 15 min[87]), much greater than a single circulation time of 15 to 30 sec. Its gradual decrease is due to inactivation in the liver.

The more complete sequence of events stimulated by the charge in plasma volume and ending the release of renin, followed by the arterial muscle contraction and raising of blood pressure, is diagramed in Fig. 10. A hemorrhage can also induce the release of renin from the kidney.[88] The "normal" blood levels of angiotensin or renin and the pressor effect vary from individual to individual,[86] being less during pregnancy and in patients with cirrhosis and ascites. In the case of hypertension there is no clear relation between blood pressure and angiotensin levels.

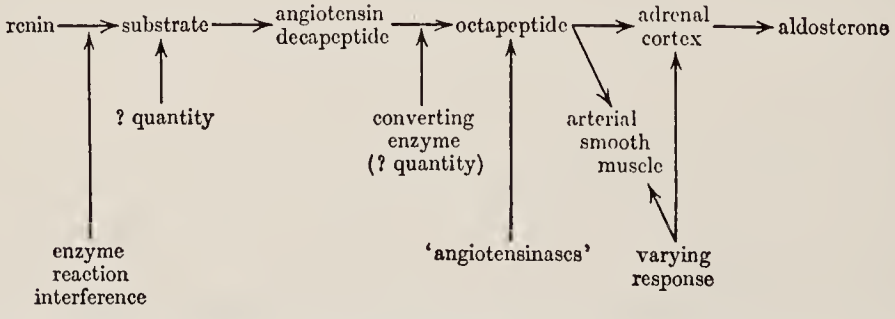


Fig. 9. The renin-angiotensin cascade as described in Ref. 86 [*Proceedings of the Royal Society* (*London*)].

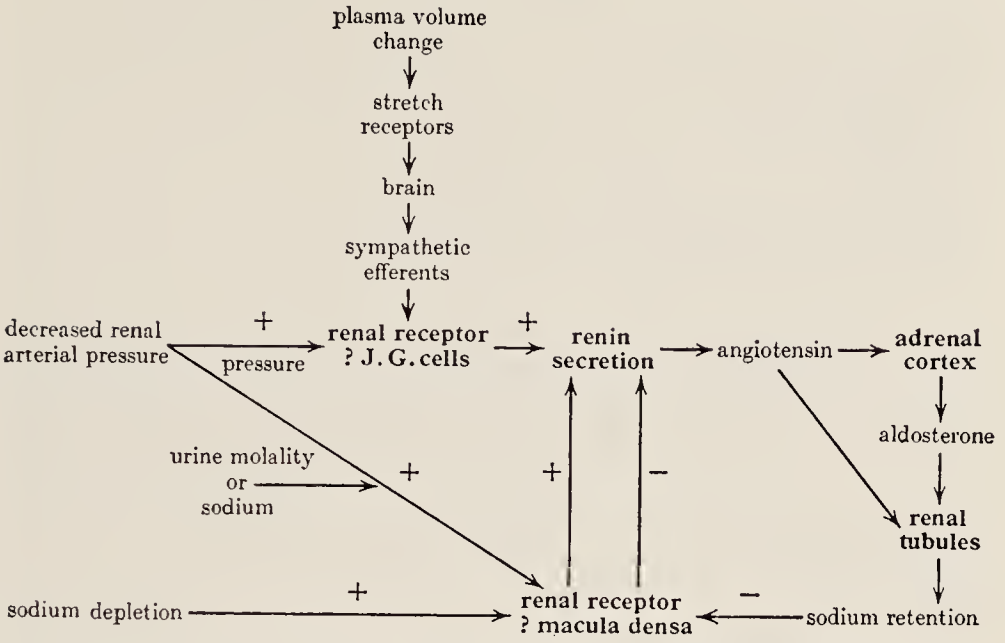


Fig. 10. The sequence of events stimulated by a change in plasma volume and leading to a release of renin followed by associated muscle contraction as diagramed in Ref. 86 [*Proceedings of the Royal Society* (*London*)].

E. The Kinin System

The concept of the kinin system has emerged from laboratory observations and not from clinical practice. The four biological effects that seem to be connected with this system are as follows:[89] (*1*) vasodilatation of blood vessels which, when induced systematically, is manifested as hypotension and (*2*) as increased permeability of walls of capillaries to plasma protein; (*3*) contraction of certain kinds of smooth muscle (called uterotomic); and (*4*) production of pain on a bare blister base by application of an appropriate plasma peptide product. All these may be induced by the end product of the plasma kinin system. The kinins are formed by enzymes, the kininogenases, acting upon substrates called kinogens, by splitting them at an appropriate point.

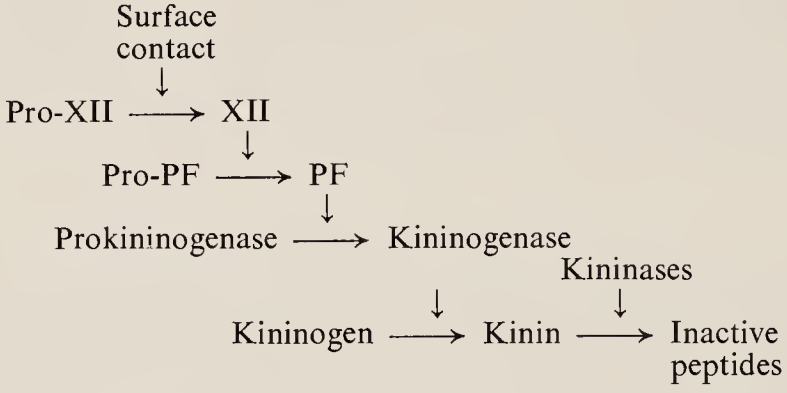
Effects (*4*) and (*2*) were observed through the production of hypotension shock in laboratory animals through application of alkaline and acid hydrolysates of proteins.[89] The modern understanding of the process was considerably advanced by Kraut, Fry, and Werle.[90] Armstrong, Dry, Keele, and Markam[91] noticed that when fluid from a burn blister was applied to the exposed base of another blister it caused marked pain. Other inflammatory exudates such as pleural effusion and the fluid from a rheumatoid arthritic joint produced a similar effect.

The amino acid sequence of three important kinins is as follows:

Arg-Pro-gly-Phe-Ser-Phe-Arg
Bradykinin[93,94]
Lys-Arg-Pro-gly-Phe-Ser-Pro-Phe-Arg
Kallidin[95]
Met-Lys-Arg-Pro-gly-Phe-Ser-Pro-Phe-Arg
Another plasma kinin

In the pure state the kinins are very active, with concentrations as low as 10^{-10} *M*. Bradykinin at this concentration causes the contraction of an isolated rat uterus and the relaxation of a rat duodenum at a concentration a factor of 10 to 100 higher. It also causes bronchoconstriction in the guinea pig and inhalation of bradykinin aerosol reduces the vital capacity in asthmatic but not in normal persons.[92]

The following scheme of a plasma cascade has been proposed by Miles.[89]



Factor XII is the Hageman factor of clotting and PF is the so-called permeability factor. Miles also suspects that certain inhibitors should be included in the diagram. For example, one of the inhibitors seems to be identical with the inhibitor of the esterase of C′1 component of complement. Incidently, with XII being common to the blood coagulation cascade, we see that there is probably some coupling between complement, kinin, and clotting.

Keele[92] has given the following list of pathological states in which the kinins may play a role: acute pancreatitis, acute gouty arthritis, rheumatic arthritis, bronchial asthma, hereditary angioneurotic edema, migraine, dumping syndrome, carcinorid syndrome, reaction to physical, chemical, or bacterial damage, and transfusion of plasma fractions. He also indicated how the Miles conjectured cascade might operate in these various conditions. We consider one of his examples and refer the reader to the original paper for an analysis of the others.

Acute attacks of gout are an inflammatory response to the deposition of sodium urate crystals in tissue. It has been shown[96] that the concentration of kinins in joint fluid is high enough to produce vasodilatation and pain. Eisen and Keele suspected that the urate crystals might trigger kinin formation in the same way that glass as a surface contact triggered the Miles cascade discussed above, that is, by activating the Hageman factor (factor XII). This conjecture was tested by first incubating synovial exudates with microcrystals of monosodium urate and applying the product to a rat uterus. The inflammation response of the uterus was the same as it would have been to a kinin concentration of 5×10^{-7} M. This action of urate crystals was prevented by hexadimethrine bromide, a substance known to prevent the activation of factor XII by a foreign surface in the blood clotting cascade.

It has already been noted above that certain factors are common to several cascades. This implies that various control mechanisms in blood plasma are related to each other. E. L. Becker[97] has examined these possible relations and sketched a possible flow chart for them, which is reproduced in Fig. 11.

III. KINETICS OF ENZYME CASCADES

The kinetics of the classical schematic form of an enzyme reaction as given by (1) is described by the rate equation

$$\frac{dP}{dt} = kE \tag{5}$$

P being the concentration of product and E that of enzyme. In this equation one postulates the amount of substrate S to be enormous or to be replaced at a rate faster than it is consumed.

The basic kinetic equations for the Michaeles-Menten Model described by (2) are

$$\frac{dS}{dt} = -k_1ES + k_{-1}C \tag{6a}$$

$$\frac{dC}{dt} = -k_{-1}C + k_1ES - k_2C \tag{6b}$$

$$\frac{dP}{dt} = k_2C \tag{6c}$$

$$\frac{dE}{dt} = -k_1SE + k_2C + k_{-1}C \tag{6d}$$

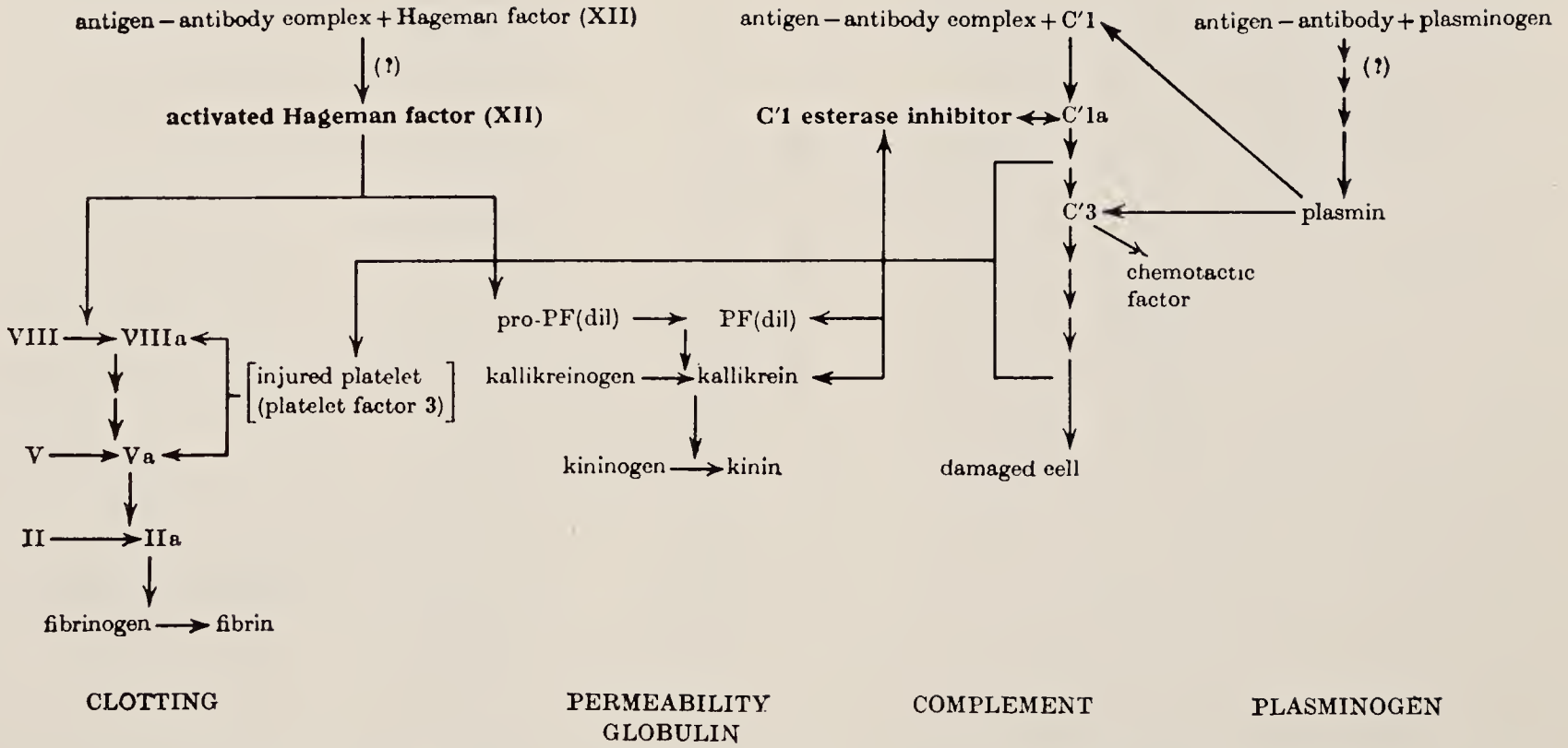


Fig. 11. A sketch of the manner in which various cascades might be interrelated, as proposed by Becker in Ref. 97, from which the flow chart has been reprinted.

the quantities S, E, C, and P representing the concentration of the components which appear in (2). From these equations one sees that $E + C =$ constant. The rate at which product is produced is, from (6c), proportional to C. Hence one has two independent equations for C and S which when solved give the product formation velocity $v = dP/dt$. The "steady-state" solutions which are found by setting $dC/dt = 0$ with $E + C = E_0$ (E_0 being the initial enzyme concentration) yields

$$v = \frac{k_1 S E_0}{S k_1 + (k_2 + k_{-1})} \tag{7}$$

the so-called Briggs-Haldane formula. No simple formula exists when the full time-dependent solutions are considered; however, many computer calculations have been made. Some of these are discussed in Refs. 16 and 17.

The above ideas can be generalized to the case of cascades defined by (3) and (4).

Consider an open cascade in which the reservoir of substrate molecules $S_1, S_2, S_3, \ldots$ is so large that they do not change their concentration with time. Then if the rate constant for formation of enzyme j from enzyme $(j - 1)$ is k_{j-1}, we have

$$\frac{dE_2}{dt} = k_1 E_1 , \qquad \frac{dE_3}{dt} = k_2 E_2 , \qquad \text{etc.} \tag{7}$$

Then for $j > 1$

$$E_j(t) = k_{j-1} \int_0^t E_{j-1}(\tau)\, d\tau \tag{8}$$

and generally,[27]

$$\frac{E_j(t)}{E_1} = \frac{t^j}{j!} \prod_{i=1}^{j} k_j \tag{9}$$

The concentration $E_n(t)$ of the final products is then

$$\frac{E_n(t)}{E_1} = \frac{(t/k)^n}{n!} \tag{10a}$$

where k is the geometric mean of the rate constants

$$k = (k_1 k_2 \cdots k_n)^{1/n} \tag{10b}$$

If n is fairly large, say, 8 to 10, as it is in the case of blood clotting, $E_j(t)$ is practically zero until $t > k$, after which it grows with great rapidity so that the chain is considered to produce an avalanche of product rather than acting as a mere amplifier.

The damped cascade exhibited in (4) is characterized by the reaction equations

$$E_i + S_i \xrightarrow{\mu_i} E_{i+1} + E_i \qquad i = 1, 2, \ldots, n \tag{11a}$$

$$E_i \xrightarrow{\lambda_{i-1}} I_i \qquad i = 2, 3, \ldots, n \tag{11b}$$

which yields the rate equations, $E_1 =$ constant and

$$\frac{dE_i}{dt} = -\lambda_{i-1}E_i + \mu_{i-1}S_{i-1}E_{i-1}\,; \qquad i = 2, \ldots, n \tag{12}$$

This is equivalent to

$$\frac{(\exp -t\lambda_{i-1})\, d(E_i \exp t\lambda_{i-1})}{dt} = \mu_{i-1}S_{i-1}E_{i-1} \tag{13}$$

which yields the recurrence formula for $j = 2, 3, \ldots, n$

$$E_j(t) = E_j(0) + \int_0^t \mu_{j-1}S_{j-1}e^{-(t-\tau)\lambda_{i-1}}E_{j-1}(\tau)\, d\tau \tag{14}$$

Again, if the initial concentration of substrates $S_1, S_2, \ldots$ is so great that essentially no change occurs in these concentrations, if $E_1 =$ constant, and if $E_j(0) = 0$ for $j > 2$,

$$\frac{E_2(t)}{E_1} = \left(\frac{\mu_1 S_1}{\lambda_1}\right)(1 - e^{-\lambda_1 t})$$

$$\frac{E_3(t)}{E_1} = \left(\frac{\mu_1 S_1}{\lambda_1}\right)\left(\frac{\mu_2 S_2}{\lambda_2}\right)\{A_1^{(2)}(1 - e^{-\lambda_1 t}) + A_2^{(3)}(1 - e^{-\lambda_2 t})\}$$

with

$$A_1^{(2)} = \frac{\lambda_2}{(\lambda_2 - \lambda_1)}\,; \qquad A_2^{(2)} = \frac{\lambda_1}{(\lambda_1 - \lambda_2)}$$

Generally it is easy to show by induction or application of theory of Laplace transforms and induction that[27]

$$\frac{E_j(t)}{E_1} = \prod_{k=1}^{j-1}\left(\frac{\mu_l S_l}{\lambda_l}\right)\sum_{k=1}^{j-1} A_k^{(j-2)}(1 - e^{-t\lambda_k})\,, \qquad j = 2, 3, \ldots, n \tag{15a}$$

where

$$A_k^{(j)} = \prod_{p=1}^{j}\frac{\lambda_p}{\lambda_p - \lambda_k}\,; \qquad \sum_{k=1}^{j-1} A_k^{(j-2)} = 1 \tag{15b}$$

Equations such as (12) were discussed many years ago by McKendrick in connection with the theory of epidemics.[98] The author has also discussed

these and other chemical rate equations, which are not necessarily linear, in Refs. 99 and 100.

One can show that as each $\lambda_j \to 0$, (15) reverts back to that obtained for the open cascade (10) (if one sets $k_j \equiv S_j\mu_j$) as it should because that corresponds to very long-lived enzymes. Two regimes can be expected to exist in the qualitative character of the solution (15) of the damped cascade rate equations. If all the characteristic relaxation times (λ_j^{-1}) are long compared with the time for E_n to reach a critical stage, one has a MacFarlane avalanche. On the other hand, if the λ's are all large, then $E_n(t)$ does not grow in such a spectacular way. Indeed, it is shown by H. C. Hemker and P. W. Hemker[27] that the damped cascade mimics the kinetics of a single enzyme step.

Hemker and Hemker[27] have also considered the Michaelis-Menten model for both the open and the damped cascades. Although the resulting formulas look more complicated, the qualitative results are essentially the same as those stated above.

The above kinetic equations of the traditional type could be expected to describe enzyme cascades of the type considered in the earlier sections. Whether, as the understanding of events in blood plasma develops further, the MacFarlane-Davie-Ratnoff open cascade or models such as Seegers' with a sensitive interplay between inhibitors and activators (or a combination of both) will become the most effective way to describe the events, the kinetic equations would have a form not unlike (12). Feedback loops could also be introduced. The equations might be more difficult to solve but their structure would not change significantly. The relative numerical values of the μ's and λ's would depend on the model. It can be expected that as the various reactants and the individual reactions are investigated experimentally, values of the rate constants will become better known.

However, rate equations such as those given above do not emphasize the control aspects of the cascades and the response to natural fluctuations. A more graphical way of displaying this feature is first to enumerate a set of states for the whole system and then to discuss the transitions from one state of the system to another. To see how this might be done consider Fig. 1, the MacFarlane cascade. If the bulk of the system in the region of interest has progressed as far as the formation of XIIa but no farther, then we say it is in state 1; if it has progressed to XIa but no farther, then the system is in state 2, etc. Then the system might be described by a ladder of states, the lowest rung representing the first state, the second rung the second state, etc. The "ground" below the ladder would represent the ground state before any stage of the cascade has developed.

Fluctuations in flow and concentrations might then excite the system to rise to various levels on the ladder only to later fall to the lower levels if

all the conditions for clotting were not met. Reaching the top (which would be level 12 in the MacFarlane cascade) would correspond to the clot being formed. When a blood vessel is damaged, the various transition probabilities (through loss of effectiveness of inhibitors for various points) for going upward from one state to the next higher would increase so that the probability of reaching to the top of the ladder in a short time is enhanced tremendously. Clearly, the more levels in the ladder the smaller the probability of the top level being reached due to fluctuations alone.

The mathematical theory of "ladder climbing" models in chemical kinetics has been discussed in Refs. 100 and 101.

Acknowledgment

The preparation of this report was conducted partly under the auspices of the APRA MRC with the University of Michigan (DAHC-15-67-0062).

References

1. E. Brand, L. J. Saidel, W. H. Goldwater, W. B. Kassell, and F. J. Ryan, *J. Am. Chem. Soc.*, **67,** 1524 (1945).
2. E. Brand, *Ann. N.Y. Acad. Sci.*, **47,** 187 (1946).
3. D. M. Kirschenbaum, *Anal. Biochem.*, **44,** 159 (1971).
4. D. M. Kirschenbaum, *Anal. Biochem.*, **49** (1972).
5. S. Moore, D. H. Spackman, and W. H. Stein, *Anal. Chem.*, **30,** 85 (1958).
6. D. H. Spackman, W. H. Stein, and S. Moore, *Anal. Chem.*, **30,** 1190 (1958).
7. F. Sanger and H. Tuppy, *Biochem. J.*, **49,** 481 (1951).
8. M. J. Dayhoff and R. V. Eck, *Atlas of Protein Sequence and Structures*, National Biochemistry Research Foundation, Silver Spring, Md., 1966.
9. J. C. Kendrew et al., *Nature*, **185,** 422 (1960).
10. M. F. Perutz, *J. Mol. Biol.*, **13,** 646 (1968); *Proc. Roy. Soc. London Ser. B*, **173,** 113 (1965).
11. *Cold Spring Harbor Symposium on Quantitative Biology*, **36,** J. Watson, Ed. (1971).
12. H. M. Sobel, *Progr. Nucleic Acid Res. Mol. Biol.*, **13A** (1972).
13. L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U.S.*, **37,** 241, 251, 729 (1951).
14. B. Jirgensons, *Rept. Progr. Phys.*, **35,** 55 (1972).
15. F. Wold, *Macromolecules, Structure and Function*, Prentice-Hall, Englewood, N.J., 1971.
16. E. A. Dawes, *Comprehensive Biochemistry*, M. Florkin and E. Stotz, Eds., Vol. 12, Elsevier, 1964, p. 88.
17. J. M. Reiner, *Comprehensive Biochemistry*, M. Florkin and E. Stotz, Eds., Vol. 12, Elsevier, 1964, p. 126.
18. E. D. Koshland, *Advan. Enzym.*, **22,** 45 (1960).
19. S. H. Koenig, R. D. Brown, and J. Studebacker, *Cold Spring Harbor Symposium on Quantitative Biology*, **36,** 551 (1971).
20. G. Navon, *Chem. Phys. Lett.*, **7,** 360 (1970).
21. J. Reuben and M. Coln, *J. Biol. Chem.*, **245,** 6539 (1970).
22. A. R. Peacocke, R. E. Richards, and B. Sheard, *Mol. Phys.*, **16,** 177 (1969).
23. A. Danchin and M. Gueron, *J. Chem. Phys.*, **53,** 3599 (1970).
24. R. W. Jones, *Biological Engineering*, H. P. Schwann, Ed., Vol. 87, McGraw-Hill, New York, 1969, p. 203.

25. R. G. MacFarlane, *Nature*, **202,** 498 (1964).
26. Levine
27. H. C. Hemker and P. W. Hemker, *Proc. Roy. Soc. London Ser. B*, **173,** 411 (1969).
28. W. E. Davie and O. D. Ratnoff, *Science*, **145,** 1310 (1964).
29. R. G. MacFarlane, *Proc. Roy. Soc. London Ser. B*, **173,** 216 (1969).
30. K. Laki, *Fibrinogen*, K. Laki, Ed., Dekker, New York, 1968, pp. 1–24.
31. A. J. Quick, *J. Biol. Chem.*, **109,** 73 (1935).
32. P. A. Owren, *Acta Med. Scand. Suppl.*, **194** (1947).
33. H. S. Kingdon, E. W. Davie, and O. D. Ratnoff, *Biochemistry*, **3,** 166 (1964).
34. R. Biggs, R. G. MacFarlane, K. W. E. Denson, and B. J. Ash, *Brit. J. Haematol.*, **11,** 276 (1965).
35. M. P. Esnouf, *Proc. Roy. Soc. London Ser. B*, **173,** 269 (1969).
36. W. H. Seegers, *Prothrombin*, Howard University Press, Cambridge, Mass., 1962.
37. F. J. Kezdy, L. Lorand, and K. D. Miller, *Biochemistry*, **4,** 2302 (1965).
38. M. P. Esnouf and F. Jobin, *Biochem. J.*, **102,** 1310 (1967).
39. P. G. Barton and P. J. Hanahan, *Biochem. Biophys. Acta*, **133,** 506 (1967).
40. J. G. G. Schoenmakers, R. Matze, C. Haanen, and F. Zilliken, *Biochem. Biophys. Acta*, **101,** 166 (1965).
41. W. H. Seegers, *Blood Clotting Enzymology*, W. H. Seegers, Ed., Academic Press, New York, 1967, pp. 1–21.
42. R. Briggs, *Proc. Roy. Soc. London Ser. B*, **173,** 277 (1969).
43. J. Roskam, *Arrest of Bleeding*, Charles C. Thomas, Springfield, Ill., 1954.
44. A. Guarder, T. Fonsen, S. Leland, A. Hellem, and P. A. Owren, *Nature*, **192,** 531 (1961).
45. G. V. R. Born, *Nature*, **194,** 927 (1962).
46. S. A. Johnson, 1967. *Blood Clotting Enzymology*, W. H. Seegers, Ed., 1967, Academic Press, New York, pp. 379–420.
47. J. F. Mustard, M. F. Glynn, T. Hovig, L. Forgenson, M. A. Packham, E. Nishizawa, and H. C. Rowrill, in *Physiology of Hemostasis and Thrombosis*, S. A. Johnson and W. H. Seegers, Eds., Charles C. Thomas, Springfield, Ill, 1967, pp. 288–326.
48. C. E. Hall and H. S. Slayter, *J. Biophys. Biochem. Cytol.*, **5,** 11 (1959).
49. E. Mihalyi, *Fibrinogen*, K. Laki, Ed., Dekker, New York, 1968, pp. 61–85.
50. G. Koppel, *Nature*, **212,** 1608 (1966).
51. G. Koppel, 1970 *Thromb. Diath. Haemorrhag. Supp.*, **39,** 70 (1970).
52. L. Lorand, 1970 *Thromb. Diath. Haemorrhag. Supp.*, **39,** 76 (1970).
53. B. Blömbäck and M. Blömbäck, *Thromb. Diath. Haemorrhag. Supp.*, **39,** 17 (1970)
54. A. G. Loewy, *Thromb. Diath. Haemorrhag. Supp.*, **39,** 103 (1970).
55. N. U. Bang, *Blood Clotting Enzymology*, W. H. Seegers, ed., Academic Press, New York, 1967, pp. 487–549.
56. J. D. Ferry and P. R. Morrison, *J. Am. Chem. Soc.*, **69,** 388, 400 (1947).
57. J. D. Ferry, M. Miller, and Shulman, *Arch. Biochem. Biophys.*, **34,** 424 (1951).
58. B. Blombäck, 1971, *Biochemical Evolution and Origin of Life*, E. Schoffernels, Ed. North-Holland, Amsterdam, 1971, pp. 112–129.
59. R. F. Doolittle and G. A. Mross, *Nature*, **643,** 225 (1970).
60. R. F. Doolittle, *Thromb. Diath. Haemorrhag. Supp.*, **39,** 17 (1970).
61. D. Merrache, *Thromb. Diath. Haemorrhag. Supp.*, **39,** 308 (1970).
62. E. A. Beck, *Thromb. Diath. Haemorrhag. Supp.*, **39,** 323 (1970).
63. M. Blombäck, B. Blombäck, and A. S. Praud, *Nature*, **218,** 134 (1968).
64. St. Magnussom, *Thromb. Diath. Haemorrhag. Supp.*, **38,** 98 (1970).

65. N. F. Rodman and R. G. Mason, *Thromb. Diath. Haemorrhag. Supp.*, **40,** 146 (1970).
66. D. J. Lyman, J. L. Brash, S. W. Chaikin, K. G. Klein, and M. Carini, *Trans. Am. Soc. Artificial Internal Organs*, **14,** 250 (1968).
67. M. L. Fucker and L. Vroman, *Proc. Soc. Exptl. Biol. Med.*, **131,** 318 (1969).
68. R. E. Baier and R. C. Dutlon, *J. Biomed. Mater. Res.*, **3,** 191 (1969).
69. G. P. McNichol, *Proc. Roy. Soc. London Ser. B.*, **173,** 285 (1969).
70. K. C. Robbins, L. Summaria, D. Ehoyn, and G. H. Barlow, *J. Biol. Chem.*, **240,** 541 (1965).
71. P. Wallén, *Abstr. Fed. Eur. Biochem. Soc.*, **107,** (1967).
72. V. Hamberg, *Proc. Roy. Soc. B.*, **173,** 293 (1969).
73. A. Engle, B. Alexander, and L. Pechet, *Biochemistry*, **5,** 1543 (1966).
74. R. G. Macfarlane and J. Pilling, *Lancet*, **1,** 10 (1937).
75. W. S. Tillett and R. L. Gardner, *J. Exptl. Med.*, **58,** 485 (1933).
76. J. Denys and H. de Marbaix, *Cellule*, **5,** 197 (1889).
77. A. A. Sharp, *Proc. Roy. Soc. London Ser. B*, **173,** 311 (1969).
78. S. Niewiarowski and E. Kowalski, *Bull. Acad. Polon. Sci. Ser. Sci. Biol.*, **5,** 169 (1957).
79. M. Seligman and V. Nussenzweig, *Proc. 8th Congr. Eur. Soc. Haem. Vienna* (1961), p. 120.
80. P. L. Mollison, *Proc. Roy. Soc. London Ser. B*, **173,** 377 (1969).
81. P. J. Lachmann, *Proc. Roy. Soc. London Ser. B*, **173,** 371 (1969).
82. P. J. Lachmann, *Clinical Aspects of Immunology*, P. G. H. Gell and R. R. A. Coombs Eds., Blackwell, London, 1968, Chap. 14.
83. H. J. Muller-Eberhard, *Advan. Immunol.*, **8,** 1 (1968).
84. H. Green, P. Barrow, and B. Goldberg, *J. Exptl. Med.*, **110,** 699 (1959).
85. C. G. Cochrane and H. J. Muller-Eberhard, *J. Exptl. Med.* **127,** 371 (1968).
86. W. S. Peart, *Proc. Roy. Soc. London Sec. B*, **173,** 317 (1969).
87. R. L. Hodge, R. D. Lowe, and J. R. Vane, *J. Physiol. London*, **185,** 613 (1966).
88. O. A. Scornik and A. C. Paladini, *Am. J. Physiol. London*, **206,** 553 (1964).
89. A. Miles, *Proc. Roy. Soc. London Sec. B*, **173,** 341 (1969).
90. H. Kraut, E. K. Frey, and E. Werle, *Hoppe-Seylers Z. Physiol. Chem.*, **222,** 73 (1943).
91. D. Armstrong, R. M. L. Dry, C. E. Keele, and J. W. Markham, *J. Physiol. London*, **117,** 4 (1952).
92. C. A. Keele, *Proc. Roy. Soc. London Sec. B*, **173,** 361 (1969).
93. D. F. Elliott, G. P. Lewis, and E. W. Horton, *Biochem. Biophys. Res. Commun.*, **3,** 87 (1960).
94. R. A. Boissonas, S. Guttmann, and P. A. Jaquenoud, *Helv. Chim. Acta*, **43,** 1349 (1960).
95. J. V. Pierce and M. E. Webster, *Biochem. Biophys. Res. Commun.*, **5,** 353 (1961).
96. K. L. Melmon, M. E. Webster, S. E. Goldfinger, and J. E. Siegmiller, *Arthritis Rheumat.*, **10,** 13 (1967).
97. E. L. Becker, *Proc. Roy. Soc. London Ser. B*, **173,** 383 (1969).
98. A. McKendrick, *Proc. London Math. Soc.* **13** (2), 401 (1914).
99. E. W. Montroll, *Lectures Theor. Phys.* (*Theor. Phys. Inst., Univ. Col.*), **X-A,** 531 (1968).
100. E. W. Montroll, *Energ. Met. Phenom.*, **3,** 123 (1967).
101. E. Montroll and K. Shuler, *Advan. Phys. Chem.*, **1,** 361 (1958).

BÉNARD CONVECTION

E. L. KOSCHMIEDER

College of Engineering and Center for Statistics and Thermodynamics, The University of Texas, Austin, Texas

CONTENTS

I. Introduction 177
II. Basic Theoretical Considerations 178
III. Basic Experimental Considerations 185
IV. Early Experiments 186
V. Heat Transfer and Temperature Measurements 189
VI. Surface Tension Effects 192
VII. The Planform and the Wavelength 194
VIII. Theory of Finite Amplitude Convection 200
IX. Advanced Experiments 203
Acknowledgments 209
References 210

I. INTRODUCTION

This review is concerned with the many consequences of two classical studies, the experiments of Henri Bénard[1] and the theory which Lord Rayleigh[2] developed in order to explain the results of Bénard's experiments. For historical reasons and for simplicity the topic of both studies is referred to here as Bénard convection. We know now that Rayleigh's theory was concerned with something slightly different from what Bénard had observed. No injustice will be done to the genius of Lord Rayleigh, though, if his name does not appear in the designation of the problem. In the time span since the publication of the classical papers the problem has expanded amazingly; a recent compilation of the literature by M. Velarde lists more than 400 relevant contributions. Surveys have also been published, the two most often cited being those by Chandrasekhar[3] and Segel.[4] Both look at the problem from the theoretical point of view. Since then Bénard convection has been studied intensively and to some extent changed considerably. This review presents a critical comparison of the experimental results with the theoretical studies; it focuses on the facts and the problems of Bénard convection.

In order to avoid misunderstandings we define precisely the scope of this review. We stick closely to results pertaining to the linear theory of convection, which is the solid basis for all convection work. Hence we discuss convection at the critical Rayleigh number as well as convection near R_c with slightly supercritical Rayleigh numbers, say, up to about $10R_c$. The modern extensions of linear theory apply to low supercritical conditions and we think that up to $10R_c$ we are still on safe ground. With linear theory goes usually the condition of shallow layers; hence convection in bounded containers in which the walls are dominant is not covered. Concerning this topic see the recent review of Ostrach.[5] Also not covered, though the object of linear analysis, are the interesting chemical instabilities and convection in two-component fluids; this problem has been reviewed by Schechter and Velarde.[6] Omitted likewise are turbulent convection, as being too uncertain with regard to theory and experiments, and the exciting geophysical implications of convection, which have been reviewed by Turcotte and Oxburgh.[7] Also not covered are convection phenomena in the atmosphere and in astrophysics. It is imperative that we first understand clearly the puzzling problems encountered in laboratory Bénard convection before we can try to explain convection as it occurs in nature with its many complexities.

Finally a note on how this review proceeds: methods, either theoretical or experimental, are usually not mentioned. There are only about 100 references, which means that not each and every paper concerning Bénard convection has been cited. It is an obligation of the author of a review to reduce the overwhelming amount of literature to manageable proportions. The selection of the cited papers has been made with the following point of view in mind. Theoretical studies have to be based on physically realistic assumptions, they have to be formally correct, and the results have to be verified experimentally. The results of experiments have to be verified by control experiments and have to be explained theoretically. The author hopes to have quoted all essential contributions to this elaborate cross-checking procedure, which in the end decides what is true.

II. BASIC THEORETICAL CONSIDERATIONS

In order to define some terms to be used repeatedly later on, we give here a brief outline of the basic theoretical concept. Linear stability theory, based on Rayleigh's[2] fundamental study, analyzes the response of a liquid layer with an unstable density distribution (cold heavy liquid on top of warm light liquid) against infinitesimal disturbances of arbitrary size (wave number). In the mathematical analysis all nonlinear terms in the Navier-Stokes equation as well as in the equation of heat conduction are

neglected. Use is also made of the Boussinesq approximation, which in this case amounts to the statement that we treat the density ρ as a constant in all terms of the equations, except where combined with gravity. The equation of state is then

$$\rho = \rho_0(1 - \alpha\,\Delta T) \tag{1}$$

It is further assumed that the physical properties of the fluid are constant. Since the kinematic viscosity ν, the thermal diffusivity κ, and the volume expansion coefficient α actually vary with temperature we assume that

$$\left|\frac{\Delta\nu}{\bar{\nu}}\right| \ll 1\,, \qquad \left|\frac{\Delta\kappa}{\bar{\kappa}}\right| \ll 1\,, \qquad \text{and} \qquad \left|\frac{\Delta\alpha}{\bar{\alpha}}\right| \ll 1 \tag{2}$$

For a rigorous analysis of the Boussinesq approximation see Mihaljan.[8] For the application of the Boussinesq approximation to a compressible ideal gas see Spiegel and Veronis.[9] Boundary conditions for the two planes which constrict the fluid layer vertically also must be considered. These boundaries are either rigid (no slip) or free (no stress). No stress on a free surface excludes surface tension. Normally no lateral boundary conditions are taken into account since the liquid layer is considered to be of infinite horizontal extent. The thermal boundary conditions are uniform temperatures on the top plate as well as on the bottom plate. It is further required that the temperature perturbation on the boundaries vanishes, which amounts to the requirement that the thermal conductivity of the top and bottom plates is excellent as compared to that of the fluid.

Using the linearized Navier-Stokes equation, linearized heat conduction equation, and the equation of continuity one obtains, after elimination of all other variables, an equation for the vertical velocity component w

$$\left\{\left(\frac{\partial}{\partial t} - \nu\Delta\right)\left(\frac{\partial}{\partial t} - \kappa\Delta\right)\Delta + \alpha g\beta\Delta_2\right\}w = 0 \tag{3}$$

with the Laplace operator Δ and with $\Delta_2 = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$. Equation (3) is solved with the product solution

$$w = w_0 f(x, y)F(z)e^{st} \tag{4}$$

The arbitrary factor w_0 shows that neither the value (the amplitude) of the vertical velocity component nor the direction (upward or downward) of the motion can be determined. The time factor $\exp(st)$ decides the stability question. The liquid layer is stable when w decreases exponentially with time ($s < 0$, dampening). If $s = 0$ we speak of marginal stability and in the case that w increases exponentially the liquid is obviously unstable. There is also the possibility of s being complex which describes oscillatory motion, a case which is referred to as overstability. However,

overstability cannot occur in buoyancy-driven convection on a resting plate, as was shown by Pellew and Southwell.[11]

From (3) follows for marginal stability ($s = 0$), when the variables are nondimensionalized, the equation

$$\{(D^2 - a^2)^3 + Ra^2\}W = 0 \tag{5}$$

with the nondimensional operator $D = d/dz$, the wave number $a = 2\pi/\lambda$ (where the wavelength λ is the horizontal extension of the disturbance, measured in units of the liquid depth), and the nondimensional Rayleigh number

$$R = \frac{\alpha g \, \Delta T d^3}{\nu\kappa} \tag{6}$$

The horizontal velocity components follow from the equation

$$\Delta_2 f(x, y) + a^2 f(x, y) = 0 \tag{7}$$

which is often referred to as the membrane equation.

The main mathematical problem posed is the solution of (5), which is a two-parametric eigenvalue problem. For a given arbitrary wave number a this equation can be solved only by a specific R if the boundary conditions are to be satisfied. Solutions have been obtained for free-free, rigid-free, and rigid-rigid boundaries by Low,[10] Pellew, and Southwell,[11] and Reid and Harris.[12] The resulting curve for marginal stability with rigid-rigid boundaries is shown in Fig. 1. Of particular importance is the minimum of the curve which defines the critical Rayleigh number R_c. The values of R_c for different boundary conditions are as follows

$$R_c = 657.5, \quad a_c = 2.22 \quad \text{(free-free)}$$
$$R_c = 1100.6, \; a_c = 2.68 \quad \text{(rigid-free)}$$
$$R_c = 1707.7, \; a_c = 3.117 \; \text{(rigid-rigid)}$$

Since, for a given liquid layer, the Rayleigh number is a nondimensional measure of the temperature difference across the layer, R_c gives the minimal temperature difference at which convective motion in the liquid will occur. At subcritical temperature differences ($\Delta T < \Delta T_c$) the unstably stratified liquid will stay at rest even though subject to infinitesimal disturbances. At $R = R_c$ or $\Delta T = \Delta T_c$ the liquid will move and exhibit patterns with the critical wave number a_c. The eigenvalue R_c is degenerate; there is an infinite number of possible patterns which solve (7) at the same value R_c. For supercritical conditions ($R > R_c$) the problem is nonunique; for a specific R the possible wave numbers fall then into the interval given by the marginal curve.

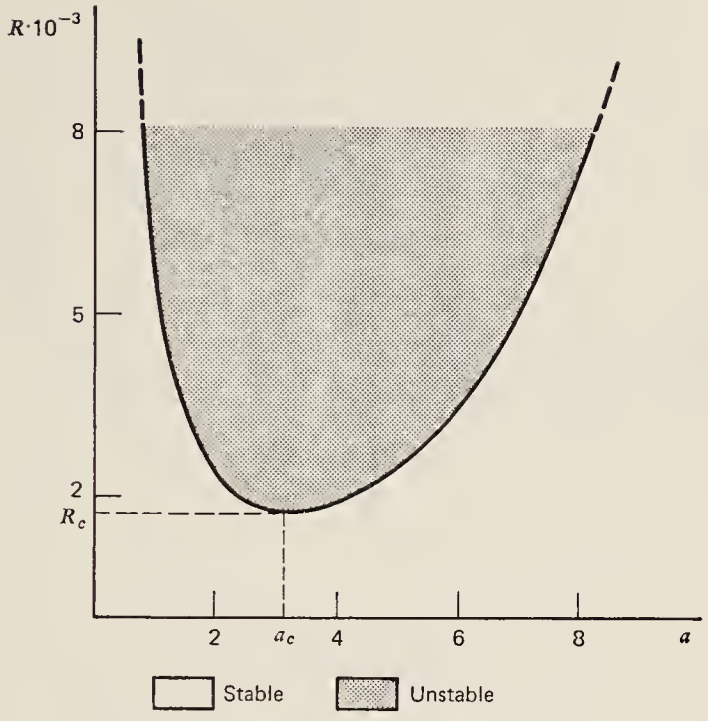
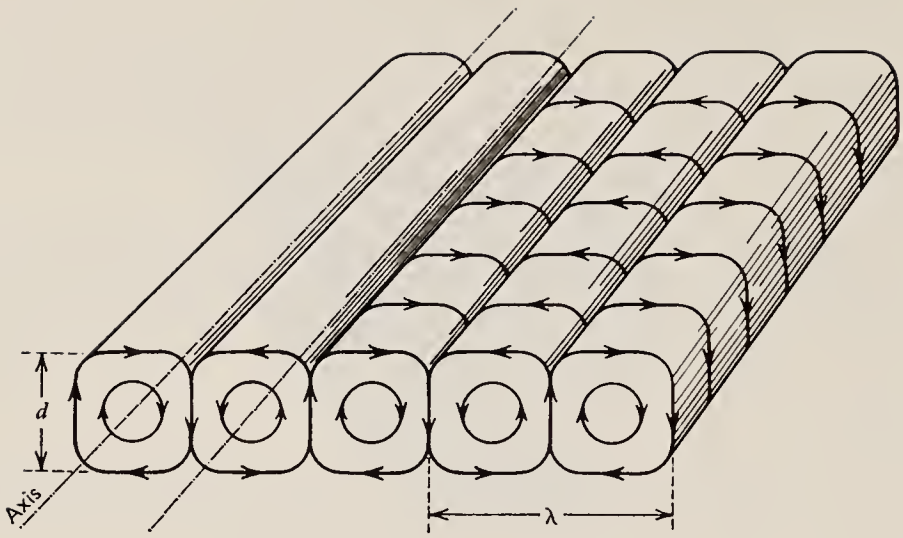
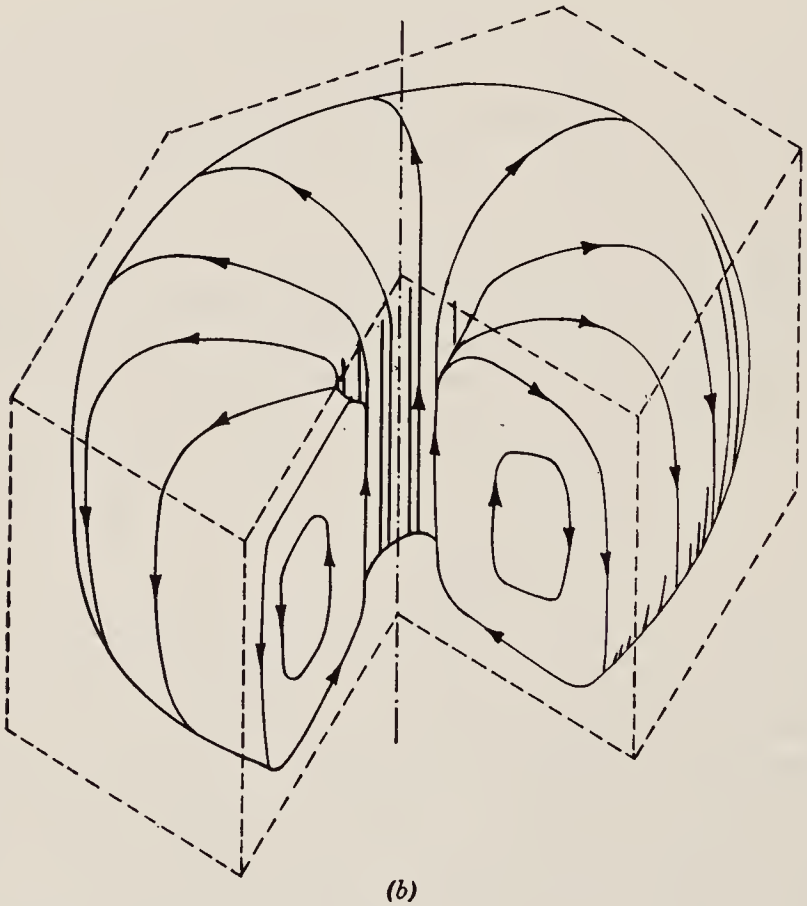


Fig. 1. Marginal curve for rigid-rigid boundary conditions.

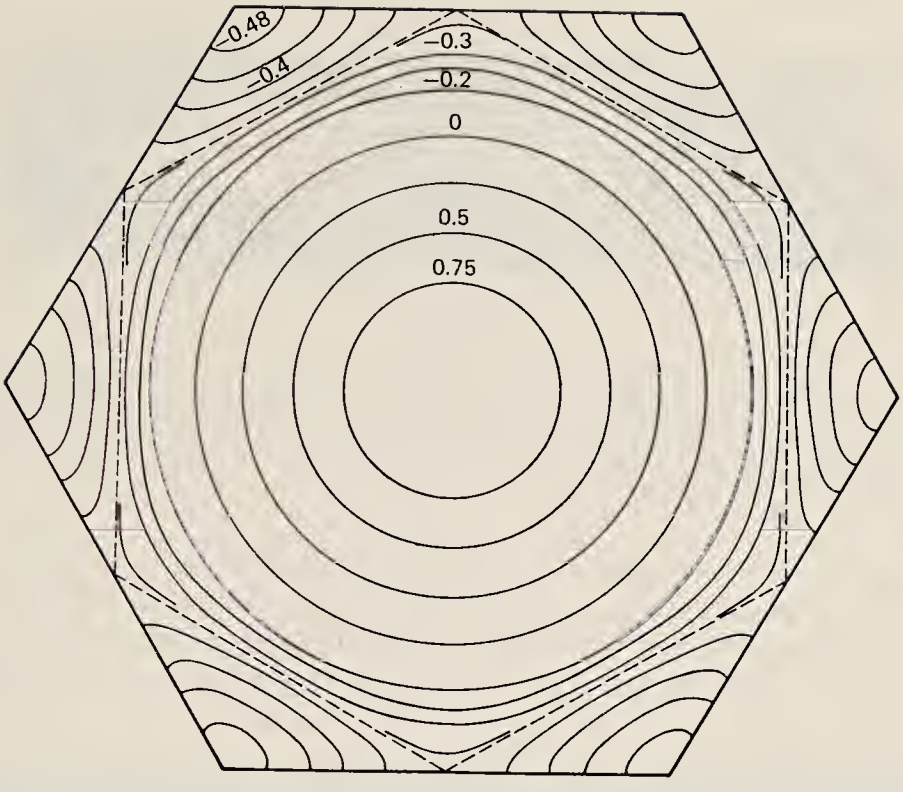
The cell patterns obtained by solving the membrane equation include straight two-dimensional rolls, equilateral triangular cells (actually a superposition of three rolls whose axes intersect each other at an angle of 60°), and hexagonal cells (composed of six triangular cells). The solution of the membrane equation for the hexagonal cells was obtained by Christopherson.[13] There are also square cells (actually a superposition of two pairs of rolls with perpendicular axes) and rectangular cells with all possible side ratios. The streamlines of rolls and hexagons are indicated in Figs. 2*a* to 2*c*. A remarkable photograph of the flow in the interior of two adjacent hexagonal cells is shown in Fig. 3. Photographs of hexagons and rolls can be seen in Figs. 4 and 8. For more detail concerning the different patterns see Chandrasekhar's book. Note, however, that the square cells as described there do not represent what one usually would call a cell, as Stuart[14] has pointed out. Chandrasekhar also discusses a generalized cell pattern. Actually, through superposition of different cells, almost any arbitrary periodic pattern can be realized mathematically. The membrane equation has also an axisymmetric solution, the ring-cell, discovered by Zierep,[15] see Fig. 2d, as well as a more general solution (Zierep[16]). Not all possible solutions of (7) are physically realistic though. Except for the square cells, no regular pattern of any rectangular cell nor a generalized



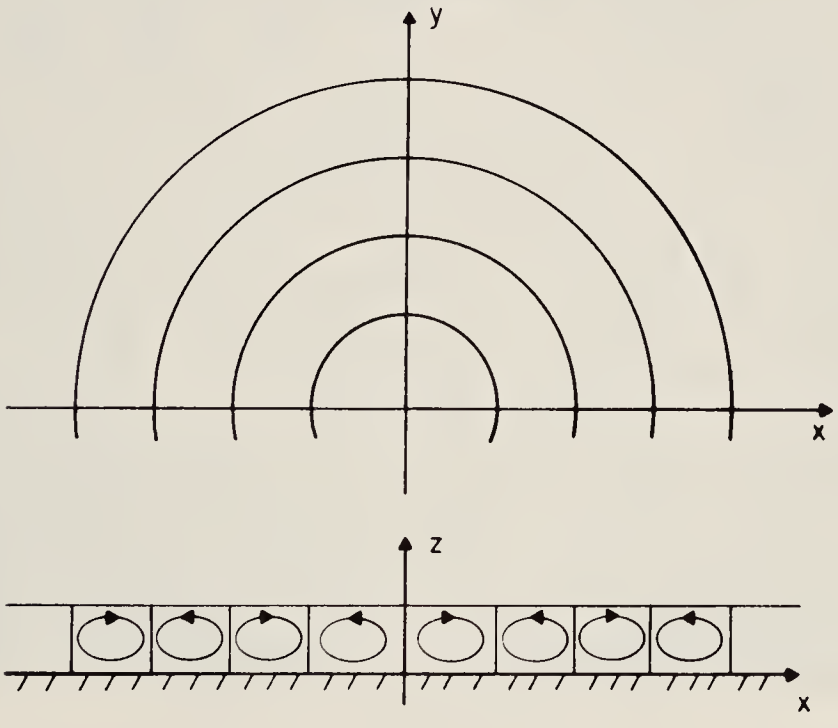
(a)



(b)



(c)



(d)

Fig. 2. (*a*) Flow in parallel rolls. Schematic. After Avsec.[104] (*b*) Flow in a hexagonal cell in a fluid. Schematic. After Avsec.[104] (*c*) Horizontal section through a hexagonal cell with lines of constant vertical velocity *w*. After Chandrasekhar.[3] (*d*) Flow in a ring-cell. Schematic. After Zierep.[15]

Fig. 3. Streak lines of the flow in the interior of two adjacent hexagonal cells. Photograph made with slit beam technique and aluminum powder as tracer. Dashed line in diagram underneath photograph shows where the light beam passed through the cells. Courtesy S. Brand, GFDL, Florida State University.

cell pattern has been observed. It remains to mention that, according to linear theory, cells which are stacked in the vertical, the so-called higher modes, are possible too. They require a supercritical Rayleigh number to be established. Higher modes, as a steady cellular pattern, have never been observed either.

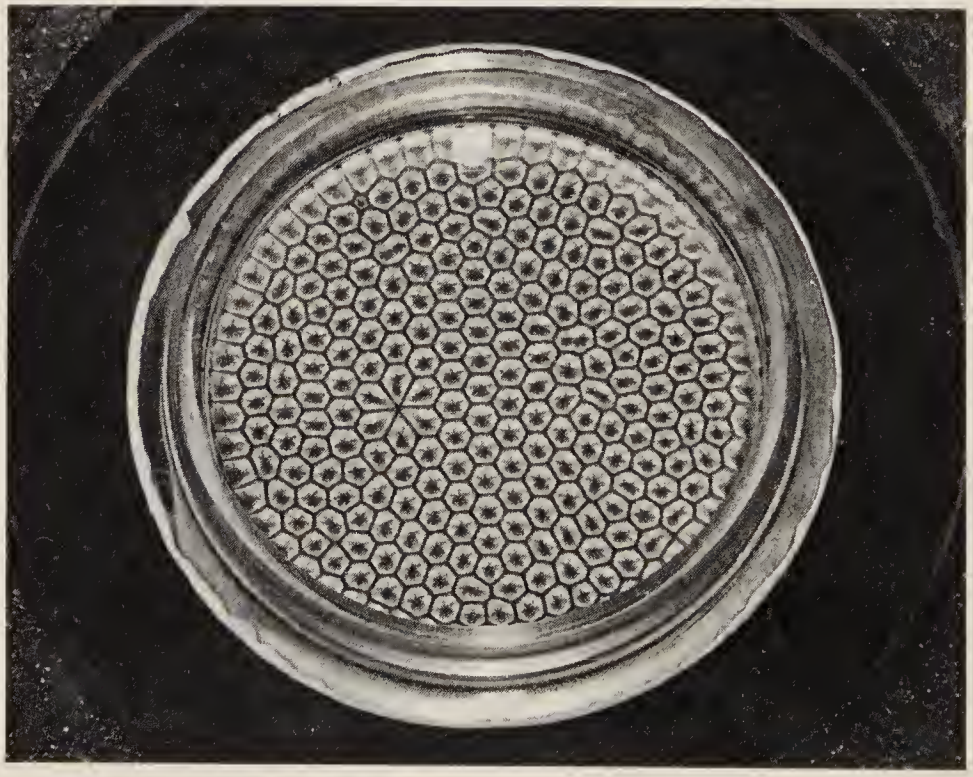


Fig. 4. Convection cells in silicone oil under an air surface. Visualization with aluminum powder. Dark lines indicate vertical motion. Bright areas indicate predominantly horizontal motion. Koschmieder, unpublished.

III. BASIC EXPERIMENTAL CONSIDERATIONS

The experimental investigation of the convective motions of a fluid on a plane plate heated from below has the following aspects: (*a*) the observation of the onset of convection and of the pattern of the motion (the planform); (*b*) the measurement of the wavelength of the motion; (*c*) the measurement of the heat flux through the convecting layer; (*d*) the measurement of the temperature distribution in the layer; and (*e*) the measurement of the velocity distribution of the motion. The aspects *a* to *e* are listed in order of increasing difficulty. The determination of the pattern seems to be trivial, yet there is still dispute about the planforms. The measurement of the wavelength is straightforward and accurate and leads to specific information, but little use has been made of it. The heat flux, being an integral quantity, is not very specific and is best used to determine the critical Rayleigh number. The measurement of the temperature distribution and the velocity distribution is still in a rudimentary state.

Repeatedly the question has been raised on how to evaluate the results of a convection experiment. This can, of course, be done only by relating the experiment to a theory to be checked. However, since all convection theories work with uniform temperatures on the top and bottom boundaries, one has to expect first of all an *entirely uniform pattern.* Nonreproducible odd shapes, irregular polygons, or wormlike rolls, for example, are a sure sign that the temperature on either top or bottom of the layer has finite imperfections. The best that experiments can do with respect to uniform temperature is to reduce the nonuniformities to the level of thermal noise (which still involves finite temperature fluctuations.) It is unlikely that even the best experiments have reduced the nonuniformities so far. However, experience shows that entirely uniform patterns can be established in liquids with viscosities $\nu \geqslant 0.1\ \mathrm{cm^2\ sec^{-1}}$. (For smaller viscosities the critical temperature differences are usually so small that they compare with the temperature fluctuations in the boundaries.) The second elementary assumption on which almost all theories are based is the assumption of excellent heat conduction in the top and bottom boundaries. This means, in practical terms, that the liquid has to be a poor thermal conductor. Experiments with water or even mercury will hardly satisfy this condition. The third assumption made almost always is that of an infinite layer. This cannot be realized physically and is best approximated by a small aspect ratio $\eta = d/h$, being the ratio of the liquid depth to the horizontal extent of the layer. Another point to look for in the evaluation of the experiments, in particular of advanced experiments, is the control experiments. Standard physical practice demands that an

apparatus has first to reproduce flawlessly some elementary results before the setup can be used to study more complicated problems. It is by no means true that this condition has been met by all convection experiments. Finally, a comment concerning the evaluation of measured quantities—it is almost trivial to note that a measured quantity is of value only if given with its experimental error.

IV. EARLY EXPERIMENTS

We should start this section with a review of Bénard's[1] classical work. He studied the convective motions in a shallow layer of a viscous liquid (molten spermaceti) on a uniformly heated metallic plate. This plate was heated with steam from boiling water; hence the temperature of the plate was 100°C. The experimental area on the plate was either circular of 10 cm diameter or square of 10 cm length. The depth of the liquid ranged from about 0.5 mm to a little over 1 mm. The upper surface of the liquid was in contact with air of room temperature. Having applied the heat to the bottom plate for a sufficiently long time, Bénard observed a steady flow regime, exhibiting a quite regular hexagonal pattern. The regularity of the pattern increased with decreasing depth of the layer, which in this particular case is equivalent to decreasing Rayleigh number. A photograph of one of Bénard's patterns is shown in Fig. 5. The fluid motions were made visible by graphite or aluminum powder added to the liquid or through various optical methods. The motion was found to ascend along the cell centers and descend along the cell walls. Closer inspection

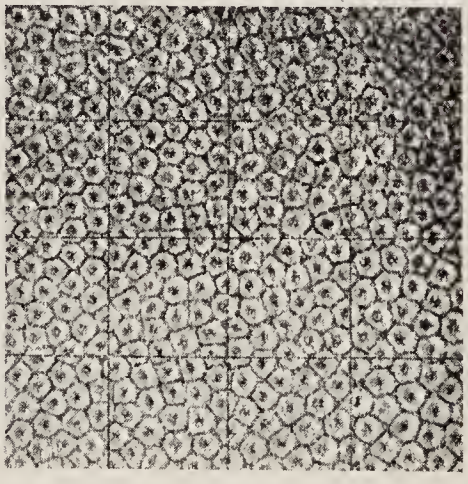


Fig. 5. Convection cells in spermaceti. One of Bénard's original photographs.

of the liquid surface revealed that the surface was depressed at the cell centers (by 1 μ or less). A considerable effort was made to determine a characteristic measure of the cell size. Photographs of the patterns containing up to 300 cells were taken in order to measure the mean distance between two cell centers. In different series the liquid depth and also the vertical temperature difference ΔT across the liquid was varied. According to Bénard (Fig. 28 of Ref. 1b) the mean cell distance has its largest value at very large temperature differences. The cell distances decrease with decreased ΔT; they reach a minimum in order to increase again slightly before the cells disappear at a certain minimal ΔT. This means, in modern terms, that after the establishment of the cells at the minimal ΔT (actually the critical temperature difference) the wavelength was observed to decrease first in order to increase substantially at larger supercritical temperature differences. (Note that Bénard was not aware of the existence of a critical temperature difference necessary for the onset of convection.) We mention finally that Bénard also made measurements of the time it takes a tracer particle to make one revolution in a closed orbit in the vertical plane within a hexagonal cell. It took around 1 sec to make one revolution within a highly supercritical cell of 1 mm depth.

It seems, in retrospect, that Bénard's experiment has confused the issue a little. His observations focused attention onto the hexagonal cells which were, for a long time, thought of as the sole representative of convective motion on a plane plate. We know now that in his experiments the air on top of the liquid introduced surface tension effects, which effects were amplified by the shallow depth of the liquid. It seems to be certain that the hexagons which Bénard observed were caused mainly by an instability due to the variation of surface tension with temperature.

About ten years after Bénard's work Dauzère[17] made convection experiments with molten wax. He observed a definite decrease in the number of cells with increased ΔT. Another paper (Dauzère[18]) showed patterns of irregular rolls and patterns with single circular cells in a field of irregular hexagonal cells. These circular cells were actually the first manifestations of the ring-cell.

In the mid 1930s a number of investigations of convective motion in gases were made. Graham[19] established the existence of irregular polygons and rolls in a layer of air heated from below. The air motions were made visible by ammonium chloride smoke. Under steady conditions the motion was *descending* in the cell centers, as opposed to *ascending* motion in the cell centers of liquids. Graham argued that the variation of viscosity ($\partial v/\partial T > 0$ in gases, $\partial v/\partial T < 0$ in liquids) was the cause for the reversal of the direction of flow. In other words, the convective motions in the liquid start where viscosity is lowest, namely, at the bottom in liquids and

at the top in gases. Graham's assumption was verified by v. Tippelskirch.[20] Chandra[21] also worked with air layers, the motion being made visible by tobacco smoke. With layers of around 1 cm in depth he observed rolls or polygonal patterns (crude hexagons) near the critical Rayleigh number. The motion in the cell centers was likewise descending. The diameters of the polygons increased with increasing ΔT. When the depth of the air layer was less than 6 mm a different type of motion, called Type II motion, occurred at subcritical Rayleigh numbers.* Since the tobacco smoke can behave quite unusually (see v. Tippelskirch[23]) one ought to be very cautious about the Type II motion. This type of motion has to be confirmed by new evidence from a carefully controlled experiment in order to be significant.

Experiments with air layers are also described in the paper by Bénard and Avsec.[24] Avsec's experiments were, however, mainly concerned with the consequences of horizontal shear. Another convection experiment with an air layer has been made by v. Tippelskirch.[23] He also established irregular rolls and polygons (see Fig. 6). He observed further that under

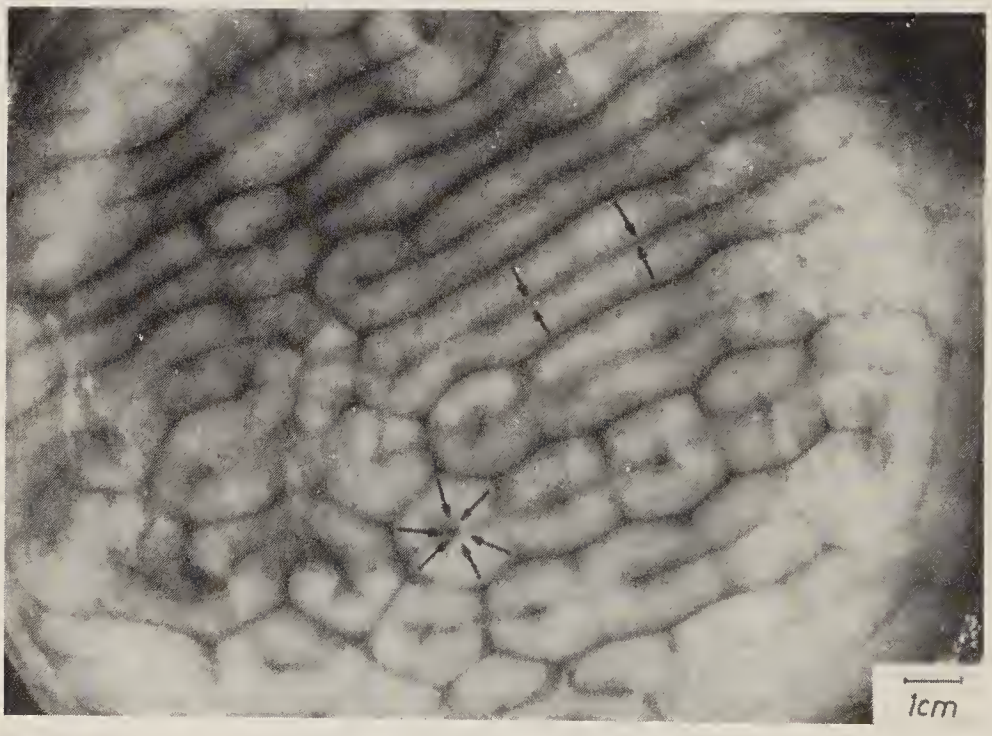


Fig. 6. Convection in an air layer. Visualization with tobacco smoke. After v. Tippelskirch.[23]

* There has been much discussion in the theoretical literature about the existence of subcritical instabilities, with some studies in favor and others flatly contradicting this possibility. The type II motion observed by Chandra has sometimes been cited as proof of the existence of such motions. For a brief review of the problem of subcritical instabilities see Davis.[22]

steady conditions a reversal of the vertical component of the motion takes place if tobacco smoke is used for visualization. This reversal seems to be due to changes in the composition of the tobacco smoke aerosol. The motion is finally descending in the cell centers. In general, there is no experiment with an air layer which is unambiguous with regard to the question of whether rolls or hexagons are the correct planform under critical conditions. Rolls are apparently more common though. An up-to-date experiment with an air layer is very desirable. A careful determination of the critical Rayleigh number in horizontal layers of gases has recently been made by Thompson and Sogin.[25] They find $R_c = 1793 \pm 80$.

V. HEAT TRANSFER AND TEMPERATURE MEASUREMENTS

Another set of experiments made in the mid 1930s was concerned with the heat transfer through a convecting fluid. Schmidt and Milverton[26] discovered that the onset of convection is accompanied by a sudden increase in the heat transfer through the liquid. Heat transfer in the resting fluid is accomplished by molecular conduction and is linearly proportional to ΔT up to the critical temperature difference ΔT_c. After the onset of convection heat is also transferred by the motion of the liquid. Therefore, a definite break in the heat transfer curve should occur at ΔT_c. Using this concept Schmidt and Milverton[26] made the very first quantitative measurement of the critical Rayleigh number for which they obtained the value $R_c = 1770 \pm 140$ (rigid-rigid boundaries). They also made an optical study of the motion in the layer between the two opaque boundaries on top and bottom. They observed that after the onset of convection a steady periodic variation of the index of refraction occurs within the liquid. This is due to horizontal variations of the vertical temperature gradient caused by cellular motions. However, the actual planform could not be determined with this method.

Schmidt and Saunders[27] essentially repeated these experiments. They observed a second break in the heat transfer curve at Rayleigh numbers of about 45,000 at which "turbulent" motions are said to begin. However, the apparatus was not sufficiently sophisticated to make this a significant result. The authors also continued the optical studies of Schmidt and Milverton and observed that the width of the cells was about twice the depth of the liquid when cellular motion was well established. This would indicate the presence of rolls. They also note that in experiments in the "turbulent" regime the index of refraction tended to vary periodically and that "the wavelength was rather greater than for the earlier stages of

cellular motion." We would now say that they observed an increase of the wavelength at large Rayleigh numbers.

Very accurate measurements of the heat transfer through a convecting liquid have been made by Silveston.[28] Five liquids with different viscosities were used. Density, viscosity, and thermal conductivity have been measured by Silveston in order to use these material constants for the determination of the Rayleigh number. The value of the critical Rayleigh number was found to be $R_c = 1700 \pm 51$. As in the experiment of Schmidt and Milverton, the critical Rayleigh number followed from the sudden increase in thermal conduction after the onset of convection. The heat transfer was furthermore measured for supercritical convection with Rayleigh numbers of up to 10^6. The heat transfer curve obtained by Silveston is shown in Fig. 7. The uncertainty in the Nusselt number was about 3%, in the Rayleigh number about 10%. Recent measurements of Pallas[29] confirm the curve shown in Fig. 7, except for the apparent "break" at $R = 3 \cdot 10^4$. Another objective of Silveston's experiments was an investigation of the relation between the heat transfer and the planform of the motion. For this purpose the top copper plate normally on the liquid layer (of 7 to 13 mm depth) was replaced by a glass plate 6 mm thick in order to be able to observe the motion. The paper does not mention that this glass plate was cooled in any systematic way other than by ambient air. Therefore, the pattern shown by Silveston will bear little resemblance to the motion which occurred in the apparatus with the cooled top copper plate in place. The initial appearance of irregular hexagonal cells which transformed in some cases at higher Rayleigh numbers into almost straight rolls (see his Figs. 6a to 6g) ought to be considered with caution. Silveston has used such patterns to plot a curve showing the wavelength increasing steadily up to Rayleigh numbers of $3 \cdot 10^4$. The values of λ he obtained are plotted in Fig. 11, where wavelengths are discussed in general.

More measurements of the heat transfer through shallow layers of silicone oil, water, and mercury have been made by Rossby.[30] He determines the heat flux by measuring the temperature drop across epoxy layers

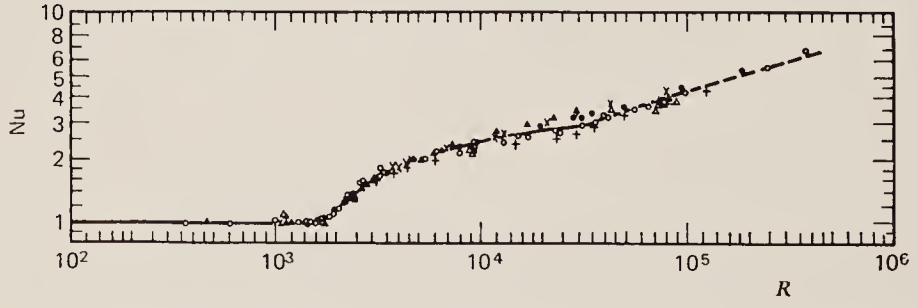


Fig. 7. Heat transfer through a convecting fluid layer. After Silveston.[28]

(thermal barriers) placed between the heat source and the bottom plate as well as between the top plate and the heat sink. Unfortunately the liquid layer is then in effect sandwiched between two thick, poorly conducting plates. Note that the thermal conductivity of water is about three, of mercury about 50 times better than that of epoxy. Rossby obtained a critical Rayleigh number $R_c = 1810$ for silicon oil, $R_c = 1760$ for water, and $R_c = 1680$ for mercury. The heat transfer curve is quite similar to Silveston's curve in the case of the oil; there are small deviations if water, and substantial deviation if mercury, was the convecting liquid.

Much numerical work has been done computing the heat transfer through convection at moderately supercritical Rayleigh numbers. A summary of the early work of Deardorff,[31] Chorin,[32] and Busse[33] was given by Schneck and Veronis.[34] Early computations assumed the convection to take place in rolls with fixed square section ($\lambda = 2$), independent of Rayleigh number. Such computations predicted Nusselt numbers too large (about 5%) in comparison with the measured values of the heat transfer (which scatter by about 3%). A similar result was obtained by Polezhaev and Vlasyuk.[35] Recent numerical studies by Lipps and Somerville[36] and Plows[37] show that the discrepancy between the numerical values and the measured values for the heat transfer vanishes if the wavelength of the motions is permitted to increase with increasing R. Numerical computations indicate a dependence of the heat flux on the Prandtl number $\sigma = \nu/\kappa$; see in particular the figure in Ref. 37, p. 60. This has to be verified experimentally. There is the further question of the dependence of the heat transfer on the type of cellular pattern. Schlüter, Lortz, and Busse[38] give a formula for the initial slope of the convective part of the heat transfer curve for different cell patterns. A measurement of the heat flux through a steady convecting liquid with an axisymmetric role pattern has been made by Pallas.[29] The initial slope that he measures for rolls for high Prandtl number fluids ($\sigma > 500$) agrees well with the slope predicted by Schlüter, Lortz, and Busse.

A limited amount of information concerning the temperature distribution in convecting liquids has been obtained in the papers of v. Tippelskirch,[20] Somerscales and Dropkin,[39] and Leontiev and Kirdyashkin,[40] for example. They made measurements with thermocouples placed into the liquid. It is somewhat uncertain how much the probes disturb the actual temperature field. However, the authors are confident that their measurements represent the temperature distribution accurately. Computed temperature fields can be found in Fromm[41] and Plows.[37] An interferometric measurement of the temperature distribution has been made by Gille.[42] This is an excellent method, introducing no disturbances into the liquid. On the other hand, the interferometric method determines only the

horizontal average of the temperature field. The measurements of Gille and the computations of Plows have established that the horizontally averaged vertical temperature distributions have curved profiles. As compared to the linear vertical temperature profile at the onset of convection, supercritical fluid layers have an increasingly isothermal midsection with strong thermal boundary layers at the top and bottom of the fluid.

VI. SURFACE TENSION EFFECTS

The period of the early experiments came to a sudden end by the discovery of Block[43] that the hexagonal cells in Bénard's experiments are almost certainly due to action of surface tension. Block removed Bénard cells in shallow hydrocarbon layers ($d \leqslant 2$ mm) instantaneously by covering the fluid with silicone monolayers. He also observed cells in liquid layers of only 50 μ depth at Rayleigh numbers a fraction of R_c. Finally he made an experiment in which the liquid was cooled from below, thus stably stratified, and yet cells appeared. Block came to the correct conclusion that the cells he observed must have been caused by the variation of surface tension with temperature. This was an essential accomplishment. Two years later Pearson[44] studied the influence of surface tension theoretically. His analysis, which neglects gravity, indicates almost conclusively that surface tension forces alone could have established Bénard's cells. The onset of convection without buoyancy ($g = 0$) is determined by the dimensionless Marangoni number

$$B = \frac{\partial S}{\partial T} \cdot \frac{\Delta T d}{\rho \nu \kappa} \tag{8}$$

S being the surface tension coefficient. The critical Marangoni number depends on a parameter L, characterizing the heat transfer from the upper surface. Note that the Marangoni number depends on d only, while the Rayleigh number depends on d^3. When in shallow fluid layers ($d < 1$) the temperature difference ΔT is increased from zero on, then the critical Marangoni number will usually be reached first. That means that in shallow fluid layers under an air surface the convective motions will in most cases be driven by surface tension.

The theoretical analysis has been continued by Scriven and Sternling[45] and Smith.[46] Scriven and Sternling neglect gravity too, but permit deformation of the free surface. They give a simple criterion for the driving mechanism of the convective motions. In steady cellular convection driven by surface tension there is upflow beneath depressions and downflow beneath elevations of the surface. They also find that there is no critical Marangoni number for the onset of stationary instability and that the

limiting case of zero wave number (meaning infinite wavelength*) is always unstable. There is no experimental evidence which supports this conclusion. However, for moderate wavelengths their results are essentially the same as those of Pearson.[44] Nield's[47] analysis is the one most suitable for experimental verification. In essence, Nield shows that convection of a liquid layer with a free upper surface (involving surface tension) heated from below (involving gravity) is determined by three parameters: the Rayleigh number, the Marangoni number, and the parameter L. Onset of convection can take place even with negative Rayleigh numbers (stable stratification) if only the Marangoni number is large enough. The wavelength of the motion at the onset of convection is noticeably larger than the critical wavelength of the rigid-rigid Rayleigh problem. Poor thermal conduction on top of the layer increases the wavelength. The planform of the motion remains undetermined. A theoretical study of Berg and Acrivos[48] indicates that surface active agents (even in trace amounts) may have very strong stabilizing effects on the onset of convection under an air surface. However, Palmer and Berg,[49] who have made some measurements of the critical Rayleigh and Marangoni numbers of a silicone oil layer, find that surface active agents that may have been present in their apparatus had no significant effect on the measured stability criteria. An attempt to determine the planform theoretically has been made by Scanlon and Segel.[50] They find that hexagonal patterns are stable solutions of their equations for Marangoni numbers ranging from just below critical to 64 times critical. This analysis works with a semiinfinite deep fluid layer, but where surface tension should not be very effective. A rigorous explanation of the establishment of the surface tension driven hexagonal pattern would be a step forward. 70 years after Bénard's experiments such a theory is overdue.

The onset of convection under an air surface has recently been studied by Koschmieder.[51] Under a thin, nearly stagnant layer of air with a uniform temperature a very regular hexagonal pattern developed in a silicone oil layer heated uniformly from below. Note that in such hexagonal cells the flow does not really correspond to the streamlines shown in Fig. 2*c*. Surface tension driven hexagons have nodes of the vertical velocity component which extend from the cell center to each of the corners. Streamlines in surface tension driven hexagonal cells have been determined by Nield.[53] They are similar to the streamlines of the flow in Fig. 3. Note also that buoyancy-driven cells under a true free surface should have an *elevated* surface over the cell centers (Jeffreys[52]), whereas the centers of cells observed under air surfaces have *depressed* cell centers, as Bénard

* Infinite wavelength means in practical terms motion with a wavelength equal to the container diameter.

showed. The critical wavelength of the convective motion under an air surface in Koschmieder's experiment was about 50% larger than the critical wavelength for rigid-rigid boundaries. This, as well as the observed dependence of the wavelength on the heat transfer on top of the layer confirms qualitatively the results of Nield's[47] study.

We finally mention a most unusual experiment concerning surface tension driven convection by Grodzka and Bannister.[54] Convection in a liquid layer with a free upper surface and heated from below has been studied in zero gravity (10^{-6} g) on board the Apollo XIV spaceship. The purpose of this experiment was to find out whether surface tension indeed causes convective motion if gravity is eliminated. Pearson's[44] theory gives the appropriate theoretical analysis. The Apollo experiment indicates that cellular convective motion occurs under these circumstances. It also indicates that a critical temperature difference has to be reached in order to start motions in the liquid layer. The second experiment made on board Apollo XVII will hopefully make the results conclusive.

VII. THE PLANFORM AND THE WAVELENGTH

Block's discovery that the hexagonal patterns observed so far in the convection experiments with fluids were due to surface tension now raised the question, which planform actually corresponds to the conditions set forth in Rayleigh's theory. Experimental evidence concerning this question was provided by Koschmieder.[55] These experiments were made with shallow circular silicone oil layers on a uniformly heated copper plate and under uniformly cooled lids. The experiments fulfill strictly the Boussinesq approximation and also the basic assumptions of Rayleigh's theory, namely, uniform temperatures on top and bottom of the liquid as well as the condition of excellent conduction at the boundaries. The aspect ratio in these experiments was 1/20 or less. For experiments in circular containers with larger aspect ratios, see Soberman[56] and Liang, Vidal, and Acrivos.[57] If in Koschmieder's experiments the fluid in the circular container was in touch with the lid (surface tension eliminated), then a pattern of circular concentric rolls appeared at the critical Rayleigh number (see Fig. 8). For an experimental confirmation of this result see Hoard, Robertson, and Acrivos.[58] For a careful theoretical analysis of the motion in circular concentric rolls see Charlson and Sani.[59] If, on the other hand, in Koschmieder's experiment the motion in the layer was restricted to a square area by a square frame, then square cells were observed. The square cells originate from a superposition of rolls which develop parallel to each side. The observed cells correspond exactly to square cells according to linear theory. It is concluded from these experiments that the

Fig. 8. Convection in a silicone oil layer heated uniformly from below. Fluid is in touch with a uniformly cooled glass lid. Just critical. Koschmieder, unpublished.

planform of the convective motion of a Boussinesq liquid on a uniformly heated plane and under a uniformly cooled lid is determined by the form of the lateral boundaries. Lateral boundary conditions remove the degeneracy of the eigenvalue R_c. This is, from a mathematical point of view, trivial and physically sound too. Lateral walls are necessarily a finite disturbance in a liquid layer with an otherwise homogeneous temperature field. Therefore, convective motion initiates at the walls and is determined by the geometry of the walls. The onset of convection at the lateral wall is shown in Fig. 9. Such motions can be maintained indefinitely if only the applied uniform vertical temperature difference is held constant. The spontaneous onset of convection in the entire fluid at R_c, which is a consequence of linear theory for fluid layers of infinite horizontal extent, never actually takes place. In shallow bounded fluid layers the critical Rayleigh number is reached when the fluid motions cover the entire plate. And only then does the heat transfer increase measurably, as Pallas[29] has shown.

We mention finally that it cannot be concluded that the appearance of circular rolls in a circular container confirms the conclusion of Schlüter, Lortz, and Busse[38] that rolls are the only stable convective motion in a

Fig. 9. Convective motions starting at the wall. Silicone oil under a glass lid. Steady state, $0.95R_c$. Koschmieder, unpublished.

Boussinesq fluid. The same mechanism that produces circular *rolls* produces also square *cells* in the square container. The straight parallel rolls of Schlüter, Lortz, and Busse can occur only on an infinite plane which cannot be realized physically.

Theoretical studies of Bénard convection in shallow liquid layers which are bounded laterally have been made by Davis[60] and Charlson and Sani.[59] Davis studies convection in rectangular boxes. He notes a very rapid increase of the critical Rayleigh number if the aspect ratio of the layer approaches the value $\eta = 1$. On the other hand, if η decreases to $\frac{1}{6}$, then the critical Rayleigh number in the box approaches rapidly the critical Rayleigh number of the infinite layer. Flow does not take place in rectangular cells but in rolls whose axes are parallel to the shorter side of the box, in agreement with observations of Koschmieder.[55] Further theoretical studies of convection in rectangular boxes have been made by Catton[61] and Davies-Jones.[62] The essence of Davis' work has been confirmed experimentally by Stork and Müller.[63] In particular, the variation of the critical Rayleigh number with the aspect ratio of the fluid layer and the appearance of rolls parallel to the short side of the box was found to be

true. Charlson and Sani[59] have studied convection in bounded cylindrical fluid layers. Again, the critical Rayleigh number increases rapidly if η approaches 1. For $\eta = \frac{1}{20}$, however, the critical Rayleigh number is found to be 1714, which differs from R_c for an infinite layer by only 0.4%, showing that an experiment with such an aspect ratio is a very reasonable approximation to an infinite layer. Flow is found to take place in circular concentric rolls, the number of rings agreeing with observations of Koschmieder.[55] It seems fair to say that the onset of Bénard convection in shallow fluid layers which are bounded laterally is sufficiently explained. The influence of lateral walls has also been analyzed by Segel.[64] This study is concerned with convection in a container whose short side measures at most twice, and whose long side at most four times the depth of the fluid. The paper of Joseph[65] on convection in containers of arbitrary shape is concerned with the stability of either the upflow or downflow in the center of the container. Note that the upflow or downflow observed experimentally by Liang, Vidal, and Acrivos,[57] to which Joseph refers, is clearly due to imperfections of the lateral insulation. It is unlikely that stability determines the sign of the motion in the center of bounded liquid layers.

Experiments made by Koschmieder[55] also provide proof of an increase of the wavelength of convective motions with increasing supercritical R and permit an accurate measurement of the wavelength. For example, it follows from Fig. 8, using the formula $\lambda = r/nd$, that the wavelength at R_c was $\lambda_c = 2.04 \pm 0.05$, the theoretical value being $\lambda_c = 2.016$. If the temperature difference across the liquid is increased slowly (quasisteady state) to supercritical values, then the wavelength increases steadily from λ_c on. The increase of λ is manifested by the consecutive disappearance of rings in the center of the plate, as shown in Fig. 10. Further evidence of an increasing λ has been provided by Rossby,[30] Krishnamurti,[66] and Willis, Deardorff, and Somerville.[67] The values given by Rossby are somewhat inaccurate since his patterns are not regular and influenced by preceding rotation of the apparatus, which is not mentioned in the text. His value $\lambda_c = 1.9$ is only 6% off the theoretical $\lambda_c = 2.016$. It is most likely that the error of measurement was at least of this magnitude. Rossby observes an increasing λ up to about $R = 2 \cdot 10^4$, where $\lambda = 2.8$. The decrease of λ that then follows, according to him, when $R > 2 \cdot 10^4$ has to be confirmed by other evidence in order to be significant. Krishnamurti[66] likewise observed an increasing λ. She discovered, furthermore, a dependence of supercritical wavelengths on the Prandtl number of the fluid. Small Prandtl number fluids show a larger increase in λ. Finally, she noted an apparent dependence of λ on the initial conditions, the wavelength showing a hysteresis. More data concerning the influence of the Prandtl number

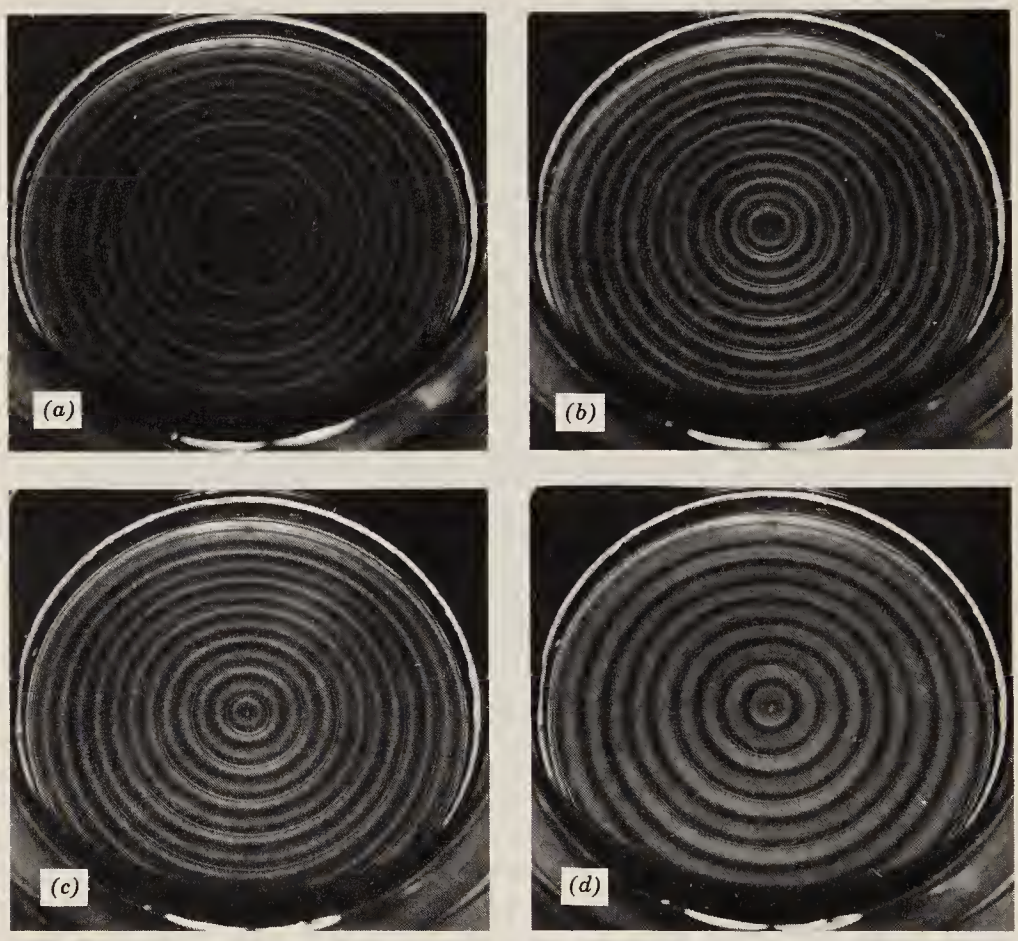


Fig. 10. The increase of the wavelength of convective motions with increasing Rayleigh number. Quasisteady state. Uniformly heated from below, uniformly cooled glass lid. Koschmieder, unpublished. (*a*) Thirteen concentric rings. Just critical. (*b*) Center ring disappearing. $R = 2.0R_c$. The fine bright circles are caused by aluminum powder settled at the bottom under the location of ascending motion. (*c*) Twelfth ring shrinking. $R = 2.88R_c$. (*d*) Nine rings left. $R = 7.23R_c$.

and of the initial conditions have been provided by Willis, Deardorff, and Somerville.[67] Some of the measured values of the wavelength are plotted in Fig. 11 for a comparison. Within the present accuracy of the measurements all experiments indicate that λ is a linear function of R up to about $10R_c$. However, the slope of the variation of λ differs for the different experiments.

Further evidence concerning the wavelength comes from another experiment of Koschmieder.[68] Under a rigid lid of very poor thermal conductivity (15 mm Lucite) the wavelength at the onset of convection

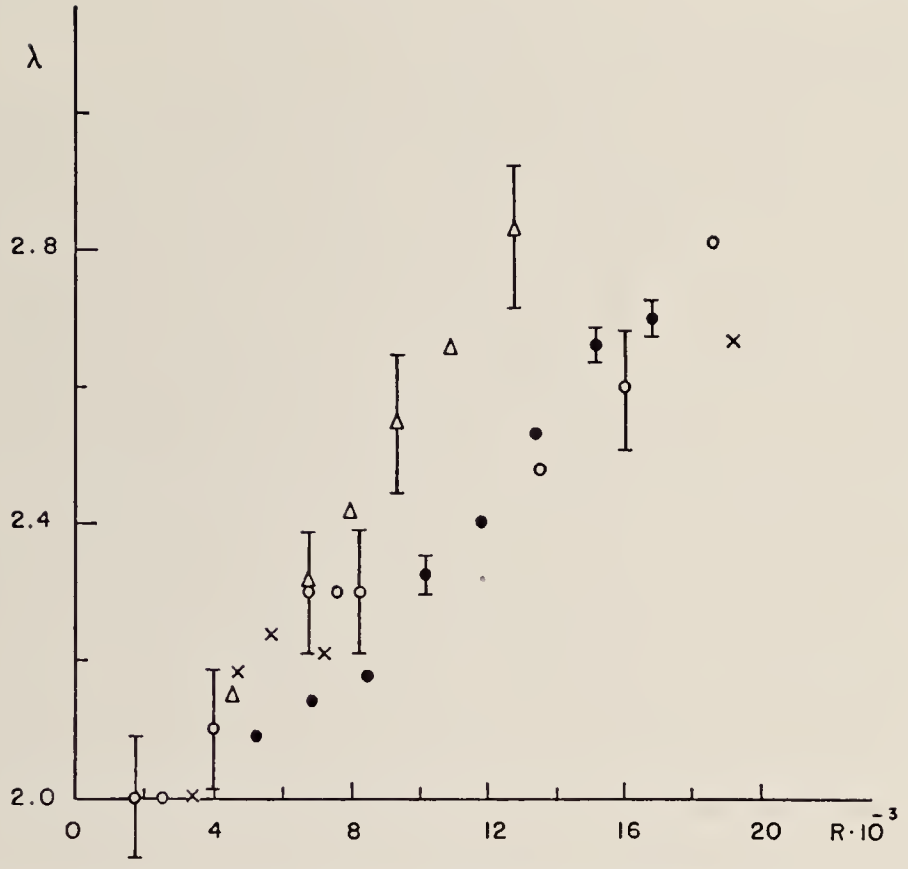


Fig. 11. Some measured quasisteady wavelengths, ΔT increasing. ○, Silveston, σ 366; ● Willis, Deardorff, and Somerville, σ 450; Δ Pallas, σ 450; × Krishnamurti, σ 860.

turned out to be *larger* than λ_c under a conducting lid. This agrees qualitatively with the analysis of Hurle, Jakeman, and Pike[69] and of Nield.[70] However, confirmation is still missing of the very large wavelengths which the theories predict for extremely poor conduction on top of the liquid layer. Koschmieder[68] also observed that supercritical temperature differences applied *suddenly* to the liquid layer between two conducting horizontal boundaries produce motions with wavelengths *shorter* than $\lambda = 2.016$. A similar observation was made at very high Rayleigh numbers by Foster.[71] Koschmieder[72] showed further, that after a sudden heating the wavelength changes slowly from its initial value $\lambda < \lambda_c$ to its steady value $\lambda > \lambda_c$ if only the suddenly applied supercritical ΔT is maintained long enough. In this case the wavelength is independent of the initial condition. The effect of initial conditions and lateral boundaries on two-dimensional convective motions has been analyzed by Foster.[73] A sudden application of the temperature difference to the liquid layer causes also a substantial

increase of the critical Rayleigh number, as observed by Koschmieder[68] and predicted theoretically by Goldstein[74] and Currie.[75]

Summarizing, we note that almost all steady convection experiments provide a (sometimes vague) indication of a λ increasing with increased Rayleigh number. Besides the specific evidence mentioned above, such indications are in Dauzère,[17] Chandra,[21] Schmidt and Saunders,[27] Chen and Whitehead,[76] and Leontiev and Kirdyashkin.[40] The sole exception is Bénard's observation of a decreasing λ near the onset of convection (surface tension and time dependency involved in his case). Yet Bénard observed increasing wavelengths too, when the applied temperature difference was much larger than the temperature difference at the onset of convection. It is, consequently, *a certainty* that *the wavelength increases with increased Rayleigh number* under steady conditions. This contradicts all theoretical analyses which predict, without exception, that *the wavelength should decrease* with increased R. Details are discussed below. In order to explain this discrepancy it has been argued that the observed increasing wavelengths might be due to poor thermal conduction in the top boundaries of the experiments. However, Koschmieder[68] and Krishnamurti[66] have shown that the increase of λ occurs also under metallic top boundaries. It has also been argued that the increase of λ might be due to effects of the lateral walls. However, the aspect ratio which Willis, Deardorff, and Somerville[67] use ($\eta = \frac{1}{8}\frac{1}{0}$) makes this argument academic. It may be true that the wavelength can be forced if the aspect ratio is small, as was indicated by Deardorff and Willis.[77] The geometry of the apparatus also might have an influence on the increase of λ. It is, for example, difficult to see how square cells in a square container can expand regularly if R is increased. Nevertheless, these are minor difficulties which can perhaps affect the rate of increase of λ. There is still work to be done to obtain an accurate curve λ versus R. The data concerning the dependence of λ on the Prandtl number as well as the dependence of λ on initial conditions likewise have to be improved in order to be certain about the latter point in particular.

VIII. THEORY OF FINITE AMPLITUDE CONVECTION

The theoretical analysis of supercritical Bénard convection is part of the current effort to develop a nonlinear stability theory. This topic has recently been reviewed by Stuart.[78] Stuart is mainly interested in the Taylor vortex problem. But all the relevant sophisticated mathematical methods with which nonlinear stability is tackled are discussed there. We restrict the comments here to work directly related to convection. The objective of the theory of finite amplitude convection is to determine the

amplitude of the motion and the pattern of the flow under supercritical conditions and to find out how a steady equilibrium state is attained after the onset of convection. Finite amplitude convection was first studied by Gorkov[79] and independently by Malkus and Veronis.[80] Malkus and Veronis present a method to determine the form and amplitude of the convective motions. They show that rectangular and hexagonal cells are finite amplitude solutions of the equations. A criterion for relative stability is introduced which selects as the realized solution the one that offers the *maximal heat transfer*. Square planforms are said to be preferred to hexagonal planforms in ordinary fluids with symmetric boundary conditions. The preferred horizontal wavelength is said to decrease with increased R. The work of Malkus and Veronis was put into rigorous mathematical form by Schlüter, Lortz, and Busse.[38] The mathematical work of Schlüter, Lortz, and Busse is considered to be impeccable. Their study has three experimentally verifiable consequences; namely, it determines (*a*) the planform (which should be rolls), (*b*) the wavelength (which should decrease as R is increased), and (*c*) the initial slope of the convective heat transfer by various cellular patterns. Item *a* cannot be verified unambiguously since Schlüter, Lortz, and Busse study motion on an infinite plane, which cannot be realized physically. Note also that straight parallel rolls introduce an arbitrary preferred direction, namely the direction of the axes of the rolls. However, the assumptions on which the theory is based do not specify any preferred direction. Point *b* is in contradiction to experimental facts, as described above. The predicted slopes in *c* have been confirmed by measurements of Pallas.[29] In addition, Schlüter, Lortz, and Busse show that at $R_c + \epsilon$ the maximal heat transfer hypothesis of Malkus and Veronis is a consequence of their theory. It apparently has never been noted that maximal heat transfer is not compatible with the balanced heat transfer which linear theory predicts to hold at R_c. To quote Chandrasekhar[3]*: "instability occurs at the minimum temperature gradient at which a *balance* can be steadily maintained between the kinetic energy dissipated by viscosity and the internal energy released by the buoyancy force." We have noted before that the available data show that one must go to longer wavelengths to match the measured heat transfer data with the computations, rather than to shorter wavelengths as the finite amplitude theory demands. With regard to this question Plows'[37] comprehensive numerical computations are decisive; his table A-5 shows that indeed more heat should be transferred by motions with wavelengths shorter than λ_c. On the contrary, however, supercritical convective motions have wavelengths longer than λ_c. This makes it certain that the maximal heat transfer hypothesis is unrealistic. On the other hand, if we assume that balanced heat transfer is maintained beyond R_c, where it is proven by linear theory,

* (l.c. pg. 34)

then it is possible to explain qualitatively the variation of the wavelength with increasing supercritical R. And at the same time the increase of λ_c under an insulating lid as well as the decrease of the wavelength after sudden heatings is understandable, as was pointed out by Koschmieder.[68] Davis[60] has shown that the balance theorem also explains qualitatively the increase of the critical Rayleigh number in bounded containers with small η.

There are many other theoretical investigations of supercritical convection. We discuss now only those which bear upon the heat transfer and the wavelength and defer to the next section the studies that investigate the consequences of the variation of viscosity with temperature. Kuo and Platzman[81] predict slopes for the convective part of the heat transfer. Herring,[82] using a quasilinear approximation, computes the heat transport and finds a wavelength decreasing with R. Platzman[83] predicts the heat transfer and decreasing wavelengths on the basis of maximal growth rates

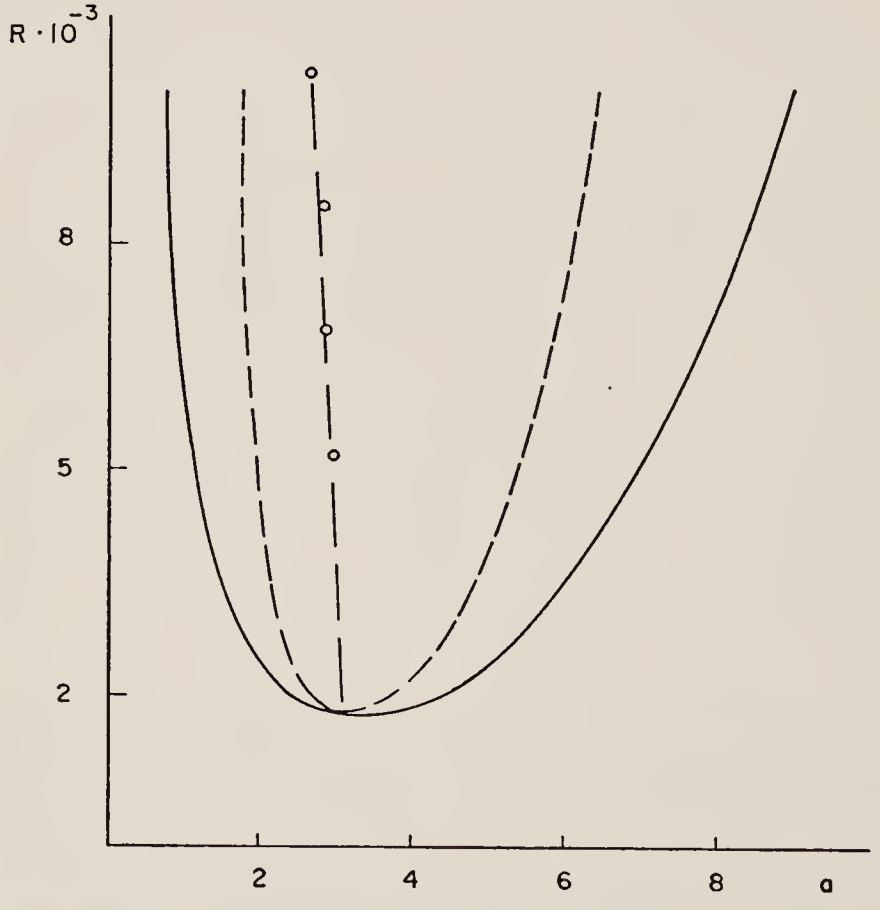


Fig. 12. Marginal curves for rigid-rigid boundaries. Solid line according to linear theory. Dashed line after Kogelman and DiPrima.[86] Circles mark measured quasi-steady state values from Willis, Deardorff, and Somerville.[67]

and maximal heat transfer. Roberts[84] determines the Nusselt number as a function of R and finds the wavelength decreasing with increasing R. Busse[33] obtains, for infinite Prandtl number fluids, a new interval for the possible wavelengths with the most stable solution having a slightly decreasing wavelength. Finally, in a recent numerical study of wave number selection in finite amplitude convection Ogura[85] finds that the wavelengths of cells which result from random initial perturbations and those which give maximum heat transfer decrease with increasing Rayleigh number. There is apparently not a single theoretical study which would predict a wavelength increasing necessarily with increased R. The latest progress in the study of nonlinear stability, concerning Bénard convection as well as Taylor vortices, comes from the paper of Kogelman and DiPrima.[86] Their analysis, which is based on the studies of Eckhaus,[87] gives a formula which permits the calculation of a new interval for the unstable wavelengths, which is shown in Fig. 12. This interval seems to be the most realistic developed so far. Yet the simple question remains, which wavelength will be observed at a given steady supercritical R, if the Rayleigh number is increased quasisteadily from R_c on. It will be an essential step forward in the theory of hydrodynamic stability when this question is answered.

IX. ADVANCED EXPERIMENTS

The experiments to be discussed in the following chapter are concerned either with effects caused by non-Boussinesq conditions or with effects caused by conditions differing otherwise from Rayleigh's theory. These experiments give more insight into the complexities of the instability or are important from the point of view of the application. The first experiment that falls under this category was made by v. Tippelskirch.[20] It provided proof that the direction of flow in the center of hexagonal cells is determined by the sign of $\partial v/\partial T$ of the respective liquid, as was surmised by Graham.[19] Using molten sulfur (whose dynamic viscosity first decreases in the range from 120 to 153°C, then increases sharply from 153 to 180°C), v. Tippelskirch demonstrated that the circulation in the cells does indeed reverse its direction if the temperature of the sulfur is increased beyond 150°C. A theoretical explanation of the dependence of the direction of the flow on $\partial v/\partial T$ was first provided by Palm,[88] a simple explanation was given by Zierep,[89] and further work was done by Palm and Oiann.[90] Another experiment of v. Tippelskirch[91] was concerned with the existence of the ring-cell. Zierep[15] had shown theoretically that a system of circular concentric rings should form around a single disturbance in an otherwise uniform, critical temperature field (see Fig. 2d). The existence of this motion was verified by v. Tippelskirch.

Much theoretical work has been done concerning the question whether a small $\partial\nu/\partial T$ might not cause hexagonal cells to be established at R_c. The hexagonal cells discussed now occur under a rigid surface and are free from influences of surface tension. This work was started by Palm[88] and then pursued by Segel and Stuart,[92] Segel,[93] Roberts,[84] Busse,[94] Palm, Ellingsen, and Gjevik,[95] and Davis and Segel.[96] It appears that this work, at least at the time of its conception, was influenced by the former belief that hexagons are the true planform of convection. The results of these studies indicate that there may be subcritical instabilities in a small range of Rayleigh numbers $R < R_c$. For small supercritical Rayleigh numbers $R_c \leqslant R \leqslant R_1$ hexagons are said to be the only stable solution, followed by an interval $R_1 < R < R_2$ in which rolls as well as hexagons are said to be stable. For all $R > R_2$ rolls should be the only stable motion. Two experiments have been made to verify the essence of these theories. The actual problem then is whether the influence of $\partial\nu/\partial T$ will be large enough to overcome the influence of the lateral walls and establish a pattern of its own. The experiment of Somerscales and Dougherty[97] is inconclusive since they do not prove that they can establish a perfect roll pattern in control experiments in which $\partial\nu/\partial T$ is negligibly small. The experiment of Hoard, Robertson, and Acrivos[58] works with a novel and decisive feature, the hydrocarbon Aroclor, the viscosity of which decreases exponentially with increased temperature. Using a 7 mm deep layer of Aroclor, Hoard, Robertson, and Acrivos first established a pattern of circular concentric rolls at a vertical temperature difference $\Delta T_c = 6.5°\text{C}$. However, in a layer only 3.64 mm deep a hexagonal pattern developed at $\Delta T_c = 18.7°\text{C}$, when $\partial\nu/\partial T$ was consequently much larger (ν at the bottom was then reduced to about 0.2 of its value on top of the layer). The hexagonal pattern so established persisted as such to about twice the critical ΔT. Though the appearance of the hexagons supports the above-mentioned theories, the persistence of the hexagons to twice ΔT_c is not in agreement with the same theories that also predict the transformation from hexagons to rolls at slightly supercritical R. One wonders, furthermore, whether the Aroclor experiment with its extraordinary large $\partial\nu/\partial T$ really relates to the theories which assume a small $\partial\nu/\partial T$.

The effects which the variation of viscosity should cause can likewise be produced by changing the mean temperature of the fluid, according to a theoretical study of Krishnamurti.[98] The objective of Krishnamurti's[99] experiment was then the verification of her theoretical results, which predict the appearance of hexagonal cells if the mean temperature is varied within certain limits. She measured the heat flux and took photographs of the convective motions between the two metallic plates by a very elegant method. The pictures show that hexagonal cells indeed

appear in the liquid at subcritical temperature differences if the mean temperature is varied.* On the other hand, with the same fluid in the same apparatus but with constant mean temperature a roll pattern appeared at the normal critical temperature difference ΔT_c. This observation, as well as the heat transfer curves measured for steady rolls and for hexagons with changing mean temperature, confirms the results of her theory. A note of caution concerning the control experiments must be added. Rolls in a square array (as in her Fig. 5a) are not a solution of linear theory when steady conditions prevail in a square container with uniform temperatures on top and bottom. Under such conditions the rolls should superpose and form square cells throughout the layer, not only along the diagonals. Also, the initial pattern should stay and should not transform slowly into a pattern of straight rolls oriented in x- or y-direction or diagonally.

Another experiment studying the onset of convection under the influence of large variations of physical properties (in this case the coefficient of thermal expansion) has been made by Caldwell.[100] Salt water is the convecting liquid, the heat flux is a given parameter, and the heat flux curves are measured. Control experiments with high temperature and pressure give results in excellent agreement with previous measurements. The critical Rayleigh number measured is $R_c = 1690 \pm 17.5$. At low temperatures (apparently at around 0°C), at which the coefficient of expansion varies considerably, quite different results were obtained. Just after the transition the heating curve goes through a minimum, instead of the usual increase of the heat transfer. There are furthermore (irregular) oscillations in ΔT just above transition if the heat flux is maintained. Finally the Rayleigh number R_c is much larger (up to 23%) than predicted. It is difficult to assess the theoretical implications of this experiment, in which the Soret effect is probably involved. The increase of the critical Rayleigh number is about 100 times larger than is to be expected from a formula for R_c of Davis[101] which takes property variations into account. The oscillations in ΔT after transition agree qualitatively with a study of Busse[102]; however, he predicts a decrease of R_c, whereas actually an increased R_c occurs. Nevertheless, Caldwell's experiment is another indication that large variations in the physical properties might have significant consequences. However, we note that, except for the work of v. Tippelskirch and its explanation, there is as yet no absolutely unambiguous experimental evidence which would correlate property variations with new effects, nor is there an unquestionable explanation of the experimental results by theoretical studies.

* These subcritical motions cannot confirm the theories which predict subcritical instabilities, since the motions in this experiment are clearly caused by the changing mean temperature.

Another method to study supercritical convection has been introduced by Chen and Whitehead.[76] They study the stability of straight parallel rolls which have been forced onto the liquid layer during a rapid heating from below by an additional radiative heating through parallel slits from above. Such experiments can produce interesting results if they are used to determine the possible range of modes of supercritical convection. Actually what Chen and Whitehead call the "neutral line for three-dimensional instability" agrees quite well with the instability interval for supercritical R which follows from Kogelman and DiPrima.[86] However, one has to keep in mind that these rolls are initially forced and may not be realized under steady conditions. It is therefore unfortunate that no description is given of the pattern that appears in their apparatus after quasisteady heating is applied from below without radiative heating from above. It also should have been mentioned that any convective pattern (e.g., hexagonal cells as well as circular rolls or cell combinations) can be produced artificially by analogous methods. Knowledge of this fact removes the "straightish rolls" of Chen and Whitehead from consideration as a possible planform of the usual convection problem. In the paper of Busse and Whitehead[103] an attempt is made to use similar experiments to verify a stability analysis of Busse.[33] Busse and Whitehead discuss two instabilities, the zigzag instability and the cross-roll instability. Both instabilities do not, to quote, "occur homogeneously throughout the layer" but rather "over regions a few wavelength in diameter." This patchy structure is a clear sign of local deficiencies of the temperature distribution. On a plate with a temperature of mathematical uniformity no point is distinguished from another. Infinitesimal disturbances as well as thermal noise, however, are present everywhere on that plate and cause motion *uniformly* over the entire plate. It would again be important to know which pattern emerges in the apparatus after quasisteady heating of the bottom plate. Concerning the stability analysis of Busse[33] and its relation to these experiments, it appears hardly possible that there is a correlation, since the very clear initial conditions in these experiments do not appear in Busse's study.

There is considerable information available concerning the influence of horizontal shear on convective motions on a plate heated uniformly from below. This problem has been studied by Graham,[19] Chandra,[21] and in particular by Avsec.[104,105] These experiments have all been made with gases. Note that in all these experiments the glass plate covering the convecting gas is cooled by ambient air only, which might explain some of the irregularities of the patterns. Aside from minor detail these experiments agree on the possibility of producing transverse rolls with small shear and longitudinal rolls with large shear. No clear-cut information is available about the point of transformation from transverse to longitudinal

rolls. Shear seems to have little influence on the onset of convection within the range of shear studied. These experiments are in qualitative agreement with the theoretical studies of Gallagher and Mercer,[106] Deardorff,[107] Gage and Reid,[108] and Lipps.[109] However, it is necessary to supplement these experiments with new, more quantitative data in order to make a more critical comparison possible.

Somewhat related to the shear experiments is a study of convection on a nonuniformly heated plane made by Koschmieder.[110] This experiment combines heating from below with the presence of a horizontal temperature gradient. This makes the problem non-Boussinesq; the horizontal density variations are now important. The patterns that develop in this experiment consist of a number of concentric rolls with alternating smaller and larger sections. The entire motion is the consequence of a superposition of an overall circulation due to the horizontal temperature gradient upon the motion due to vertical instability. The stability of the rolls of smaller section ("counter rolls") is markedly different from that of the rolls of larger section. When the shear of the circulation caused by the horizontal temperature difference becomes too large, the rolls orient themselves in radial direction. A theoretical analysis verifying the existence of the counter rolls has been made by Müller.[111] A theoretical study of the consequences of a horizontal temperature gradient on the onset of convection has been made by Unny and Niessen,[112] showing a steady decrease of the critical Rayleigh number with increasing horizontal temperature difference.

An entire chapter in Chandrasekhar's book is devoted to the problem of convection on a uniformly heated plate under the influence of rotation. According to Chandrasekhar[3] rotation has an inhibiting effect on the onset of convection. The inhibiting effect of rotation is qualitatively explained by the consideration that the particle motions are now subject to the Coriolis force. This makes the particles move in longer trajectories in which they have to overcome more dissipation than in the nonrotating case. From the theoretical analysis it follows that the critical Rayleigh number increases with increasing Taylor number $N_T = 4\Omega^2 d^4/\nu^2$, whereas the critical wavelength decreases markedly with large N_T. With small Prandtl numbers the motion should start as overstability. Later theoretical work was done by Müller,[113] Veronis,[114,115] Küppers and Lortz,[116] and Somerville.[117] Three experiments have been concerned with this topic. The apparatus used by Nakagawa and Frenzen[118] is crude. There is, in all likelihood, no uniform steady linear temperature gradient in the liquid layer, in particular in the free surface experiments. Using deep layers of water (up to 18 cm) very large Rayleigh and Taylor numbers are obtained though the applied temperature differences and rotation rates are small.

The onset of convection has been studied for Taylor numbers ranging from 10^6 to 10^{11}. The observed critical Rayleigh numbers then range from 10^5 to 10^8 in agreement with the theoretically predicted values. One wonders, though, whether a linear stability analysis is still valid when Taylor and Rayleigh numbers have such extraordinarily large values.

The experiment of Koschmieder[119] worked with shallow silicone oil layers at relatively low Taylor numbers ($N_T \leqslant 80$). Axisymmetric concentric rolls were established in a circular container, connecting the results of the rotating experiment smoothly with the results obtained on a resting plate. The concentric rolls were of alternating smaller and larger sections. This is caused by a superposition of an overall centrifugal circulation with the motion due to vertical instability. The clear-cut influence of the centrifugal force does not quite agree with the theories describing convection with rotation. All these theories neglect the centrifugal force and study the influence of the Coriolis force alone. They assume that the fluid is at rest with respect to the plates (solid body rotation) before the onset of convection, while actually any rotating stratified fluid will set up a centrifugal circulation, forcing heavy cold liquid outward and warm light liquid inward, as in a centrifuge. The first study to consider the influence of the centrifugal force has recently been made by Homsy and Hudson.[120]

Rotating his apparatus Rossby[30] measured the heat transfer in water and mercury for Taylor numbers as well as Rayleigh numbers ranging from 10^3 to 10^7. When water was used a subcritical instability at Taylor numbers greater than $5 \cdot 10^4$ was found. With mercury there was also a subcritical instability for Taylor numbers $< 1.8 \cdot 10^4$. The latter instability is said to be due to a finite amplitude instability studied by Veronis.[114] The wavelengths of the motions in a silicone oil layer in the rotating system have also been determined for Taylor numbers ranging from 10^2 to $5 \cdot 10^4$. Flow was found to arrange itself in rolls oriented radially if the apparatus rotated rapidly. The values of λ obtained indicate that the wavelength decreases with increasing Taylor number, in agreement with the prediction of linear theories. It should be noted though, that the patterns from which these λ were determined are not regular at low and quite irregular at high Taylor numbers. There are consequently substantial errors in measurement.

Convection caused by internal heating of a horizontal fluid layer has been investigated experimentally by Tritton and Zarraga[121] and by Schwiderski and Schwab.[122] The internal heating is produced by an alternating electric current passing through an aqueous zinc sulfate solution. No heat has to be supplied from the bottom plate, but the fluid layer is cooled from the top. Both experiments observed that the onset of convection takes place in polygonal (hexagonal) form with descending motion in the cell

centers. The direction of motion in the cells is probably caused by the asymmetry between heating and cooling. Secondly, it was observed that the size of the cells was very large at large R, the radius of the cells reaching five times the depth of the layer. Finally at large R a conversion to a roll pattern was apparent. Quantitative measurements of the horizontal average of the temperature distribution have been made by Kulacki and Goldstein.[123] As in the convection experiments with fluid layers heated from below, Kulacki and Goldstein find an increasingly isothermal midsection at high Rayleigh numbers in fluid layers with internal heating. Internal heating has been studied theoretically by Roberts.[124] The theoretical analysis is beset by difficulties similar to those which we have noted earlier with the standard Bénard convection problem. According to Roberts one has to expect onset of convection in the form of rolls followed by a conversion to hexagons with downward motion in the center at substantially larger R. The wavelength of the motion should decrease gradually with increased R. The latter result has been obtained again by Thirlby.[125] Yet actually the wavelength increases very much, and onset of convection definitely takes place in the form of hexagons. Some correlation between the direction of the electric current and the orientation of the rolls at larger R has been explained by Schwiderski.[126] Schwiderski[127] has recently published a theory studying convection in internally heated fluids with new mathematical methods. It appears that some aspects of this work need clarification.

We end this article on an optimistic note by mentioning the absolutely new approach to convection experiments that has been introduced recently by Ahlers.[128] He measured the heat transfer in a shallow layer of helium heated from below at temperatures between 2 and 5°K. He found up to $150R_c$ only one observable singularity in the heat transfer, namely, the one at R_c. Measurements made at such low temperatures are vastly superior in accuracy to anything possible at room temperature. Besides the total absence of convective heat losses on the outside of the apparatus (there is a vacuum) and the almost vanishing heat capacities in the top and bottom boundaries at these temperatures, there is the fundamental advantage that the thermal noise in the top and bottom plates is reduced to near zero. In such experiments we finally arrive at the infinitesimal disturbances the theories have postulated all along.

Acknowledgments

The author would like to acknowledge the many discussions he has had over the years with his colleagues. This report is actually the outcome of a joint effort to understand Bénard convection. During the last two years in particular the author appreciated the encouragement and help of Professor I. Prigogine, Professor R. S. Schechter, and

Professor M. Velarde. Part of this work was done while the author was supported through NSF Grant GA-11484.

References

1. (*a*) H. Bénard, *Rev. Gen. Sci. Pure Appl.*, **11,** 1261–1271, 1309–1328 (1900). (*b*) *Ann. Chim. Phys.*, **23,** 62–144 (1901).
2. Lord Rayleigh, *Phil. Mag.*, **32,** 529–546 (1916).
3. S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability*, Oxford University Press, Oxford, 1961.
4. L. A. Segel, in *Non-Equilibrium Thermodynamics, Variational Techniques and Stability*, Donnelly, Herman, and Prigogine, Eds., University of Chicago Press, Chicago, Ill., 1966.
5. S. Ostrach, *Advan. Heat Transfer*, **8,** 161–227 (1972).
6. R. S. Schechter and M. Velarde, *Advan. Chem. Physics*, **26** (1973).
7. D. L. Turcotte and E. R. Oxburgh, *Ann. Rev. Fluid Mech.*, **4,** 33–68 (1972).
8. J. M. Mihaljan, *Astrophys. J.*, **136,** 1126–1133 (1962).
9. E. A. Spiegel and G. Veronis, *Astrophys. J.*, **131,** 442–447 (1960).
10. A. R. Low, *Proc. Roy. Soc. London Ser. A*, **125,** 180–195 (1929).
11. A. Pellew and R. V. Southwell, *Proc. Roy. Soc. London Ser. A*, **176,** 312–343 (1940).
12. W. H. Reid and D. L. Harris, *Phys. Fluids*, **1,** 102–110 (1958).
13. D. G. Christopherson, *Quart. J. Math.*, **11,** 63–65 (1940).
14. J. T. Stuart, *J. Fluid Mech.*, **18,** 481–498 (1964).
15. J. Zierep, *Beitr. Phys. Atmos.*, **30,** 215–222 (1958).
16. J. Zierep, *Beitr. Phys. Atmos.*, **32,** 23–33 (1959).
17. C. Dauzere, *Compt. Rend.*, **156,** 218–220 (1913).
18. C. Dauzere, *Compt. Rend.*, **156,** 1228–1230 (1913).
19. A. Graham, *Phil. Trans. Roy. Soc. London Ser. A*, **232,** 285–296 (1933).
20. H. v. Tippelskirch, *Beitr. Phys. Atmos.*, **29,** 37–54 (1956).
21. K. Chandra, *Proc. Roy. Soc. A*, **164,** 231–242 (1938).
22. S. H. Davis, in *Instability of Continuous Systems*, Leipholz, Ed., Springer Verlag, 1971.
23. H. v. Tippelskirch, *Beitr. Phys. Atmos.*, **29,** 219–233 (1956).
24. H. Bénard and D. Avsec, *J. Phys. Radium*, **9,** 486–500 (1938).
25. H. A. Thompson and H. H. Sogin, *J. Fluid Mech.*, **24,** 451–479 (1966).
26. R. J. Schmidt and S. W. Milverton, *Proc. Roy. Soc. London Ser. A*, **152,** 586–594 (1935).
27. R. J. Schmidt and O. A. Saunders, *Proc. Roy. Soc. London Ser. A*, **165,** 216–228 (1938).
28. P. L. Silveston, *Forsch. Ing. Wes.*, **24,** 29–32, 59–69 (1958).
29. S. G. Pallas, Ph.D. thesis, University of Texas, Austin, Texas, 1972.
30. H. T. Rossby, *J. Fluid Mech.*, **36,** 309–335 (1969).
31. J. W. Deardorff, *J. Atmos. Sci.*, **21,** 419–438 (1964); **22,** 419–435 (1965).
32. A. J. Chorin, Ph.D. thesis, New York University, New York, 1966.
33. F. H. Busse, *J. Math. Phys.*, **46,** 140–150 (1967).
34. P. Schneck and G. Veronis, *Phys. Fluids*, **10,** 927–930 (1967).
35. V. T. Polezhaev and M. P. Vlasyuk, *Soviet Phys "Doklady" English Transl.*, **15,** 1093–1096 (1971).
36. F. B. Lipps and R. C. J. Somerville, *Phys. Fluids*, **14,** 759–765 (1971).
37. W. H. Plows, Ph.D. thesis, University of California, Livermore, Cal., 1971.

38. A. Schlüter, D. Lortz, and F. Busse, *J. Fluid Mech.*, **23,** 129–144 (1965).
39. E. F. C. Somerscales and D. Dropkin, *Intern. J. Heat Mass Transfer*, **9,** 1189–1204 (1966).
40. H. Leontiev and A. G. Kirdyashkin, *Intern. J. Heat Mass Transfer*, **11,** 1461–1466 (1968).
41. J. E. Fromm, *Phys. Fluids*, **8,** 1757–1769 (1965).
42. J. Gille, *J. Fluid Mech.*, **30,** 371–384 (1967).
43. M. J. Block, *Nature*, **178,** 650–651 (1956).
44. J. R. Pearson, *J. Fluid Mech.*, **4,** 489–500 (1958).
45. L. E. Scriven and C. V. Sternling, *J. Fluid Mech.*, **19,** 321–340 (1964).
46. K. A. Smith, *J. Fluid Mech.*, **24,** 401–414 (1966).
47. D. A. Nield, *J. Fluid Mech.*, **19,** 341–352 (1964).
48. J. C. Berg and A. Acrivos, *Chem. Eng. Sci.*, **20,** 737–746 (1965).
49. H. J. Palmer and J. C. Berg, *J. Fluid Mech.*, **47,** 779–787 (1971).
50. J. W. Scanlon and L. A. Segel, *J. Fluid Mech.*, **30,** 149–162 (1967).
51. E. L. Koschmieder, *J. Fluid Mech.*, **30,** 9–15 (1967).
52. H. Jeffreys, *Quart. J. Mech. Appl. Math.*, **4,** 283–288 (1951).
53. D. A. Nield, *Z. Angew. Math. Phys.*, **17,** 226–232 (1966).
54. P. G. Grodzka and T. C. Bannister, *Science*, **176,** 506–508 (1972).
55. E. L. Koschmieder, *Beitr. Phys. Atmos.*, **39,** 1–11 (1966).
56. R. K. Soberman, *J. Appl. Phys.*, **29,** 872–873 (1958).
57. S. F. Liang, A. Vidal, and A. Acrivos, *J. Fluid Mech.*, **36,** 239–256 (1969).
58. C. O. Hoard, C. R. Robertson, and A. Acrivos, *Intern. J. Heat Mass Transfer*, **13,** 849–856 (1970).
59. G. S. Charlson and R. L. Sani, *Intern. J. Heat Mass Transfer*, **13,** 1479–1496 (1970).
60. S. H. Davis, *J. Fluid Mech.*, **30,** 465–478 (1967).
61. I. Catton, *J. Heat Transfer*, **92,** 186–188 (1970).
62. R. P. Davies-Jones, *J. Fluid Mech.*, **44,** 695–704 (1971).
63. K. Stork and U. Müller, *J. Fluid Mech.*, **54,** 599–611 (1972).
64. L. A. Segel, *J. Fluid Mech.*, **38,** 203–224 (1969).
65. D. D. Joseph, *J. Fluid Mech.*, **47,** 257–282 (1971).
66. R. Krishnamurti, *J. Fluid Mech.*, **42,** 295–307 (1970).
67. G. E. Willis, J. W. Deardorff, and R. C. J. Somerville, *J. Fluid Mech.*, **54,** 351–367 (1972).
68. E. L. Koschmieder, *J. Fluid Mech.*, **35,** 527–530 (1969).
69. D. T. Hurle, E. Jakeman, and E. R. Pike, *Proc. Roy. Soc. London Ser. A*, **296,** 469–475 (1967).
70. D. A. Nield, *J. Fluid Mech.*, **32,** 393–398 (1968).
71. T. D. Foster, *Phys. Fluids*, **8,** 1770–1774 (1965).
72. E. L. Koschmieder, *Bull. Am. Phys. Soc.*, **16,** 1309 (1971).
73. T. D. Foster, *J. Fluid Mech.*, **37,** 81–94 (1969).
74. A. W. Goldstein, *NASA Tech. Rept.*, **R4** (1959).
75. J. G. Currie, *J. Fluid Mech.*, **29,** 337–347 (1967).
76. M. M. Chen and J. A. Whitehead, *J. Fluid Mech.*, **31,** 1–15 (1968).
77. J. W. Deardorff and G. E. Willis, *J. Fluid Mech.*, **23,** 337–353 (1965).
78. J. T. Stuart, *Ann. Rev. Fluid Mech.*, **3,** 347–370 (1971).
79. L. P. Gorkov, *Soviet Phys. JETP*, **6,** 311–315 (1958).
80. W. V. R. Malkus and G. Veronis, *J. Fluid Mech.*, **4,** 225–260 (1958).
81. H. L. Kuo and G. W. Platzman, *Beitr. Phys. Atmos.*, **33,** 137–168 (1961).

82. J. R. Herring, *J. Atmos. Sci.*, **21,** 277–290 (1964).
83. G. W. Platzman, *J. Fluid Mech.*, **23,** 481–510 (1965).
84. P. H. Roberts, in *Non-equilibrium Thermodynamics, Variational Techniques and Stability*, Donnelly, Herman and Prigogine, Eds., University of Chicago Press, Chicago, Ill., 1966.
85. Y. Ogura, *J. Atmos. Sci.*, **28,** 709–717 (1971).
86. S. Kogelman and R. C. DiPrima, *Phys. Fluids*, **13,** 1–11 (1970).
87. W. Eckhaus, *Studies in Non-linear Stability Theory*, Springer Verlag, 1965.
88. E. Palm, *J. Fluid Mech.*, **8,** 183–192 (1960).
89. J. Zierep, *Beitr. Phys. Atmos.*, **31,** 31–39 (1958).
90. E. Palm and H. Olann, *J. Fluid Mech.*, **19,** 353–365 (1964).
91. H. v. Tippelskirch, *Beitr. Phys. Atmos.*, **32,** 2–22 (1959).
92. L. A. Segel and J. T. Stuart, *J. Fluid Mech.*, **13,** 289–306 (1962).
93. L. A. Segel, *J. Fluid Mech.*, **21,** 359–384 (1965).
94. F. Busse, *J. Fluid Mech.*, **30,** 625–649 (1967).
95. E. Palm, T. Ellingsen, and B. Gjevik, *J. Fluid Mech.*, **30,** 651–661 (1967).
96. S. H. Davis and L. A. Segel, *Phys. Fluids*, **11,** 470–476 (1968).
97. E. F. C. Somerscales and T. S. Dougherty, *J. Fluid Mech.*, **42,** 755–767 (1970).
98. R. Krishnamutri, *J. Fluid Mech* , **33,** 445–455 (1968).
99. R. Krishnamurti, *J. Fluid Mech.*, **33,** 457–463 (1968).
100. D. R. Caldwell, *J. Fluid Mech.*, **42,** 161–175 (1970).
101. S. H. Davis, Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, N.Y., 1964.
102. F. H. Busse, *J. Fluid Mech.*, **28,** 223–239 (1967).
103. F. H. Busse and J. A. Whitehead, *J. Fluid Mech.*, **46,** 801–812 (1971).
104. D. Avsec, Ph.D. thesis, L'Université de Paris, 1939.
105. D. Avsec, *Publ. Sci. Tech. Min. Air*, **155** (1939).
106. A. P. Gallagher and A. McD. Mercer, *Proc. Roy. Soc. London Ser. A*, **286,** 117–128 (1965).
107. J. W. Deardorff, *Phys. Fluids*, **8,** 1027–1030 (1965).
108. K. S. Gage and W. H. Reid, *J. Fluid Mech.*, **33,** 21–32 (1968).
109. F. B. Lipps, *J. Atmos. Sci.*, **28,** 3–19 (1971).
110. E. L. Koschmieder, *Beitr. Phys. Atmos.*, **39,** 208–216 (1966).
111. U. Müller, *Beitr. Phys. Atmos.*, **39,** 217–234 (1966).
112. T. E. Unny and P. Niessen, *J. Appl. Mech.*, **36,** 121–123 (1969).
113. U. Müller, *Beitr. Phys. Atmos.*, **38,** 1–8 (1965).
114. G. Veronis, *J. Fluid Mech.*, **24,** 545–554 (1966).
115. G. Veronis, *J. Fluid Mech.*, **31,** 113–139 (1968).
116. G. Küppers and D. Lortz, *J. Fluid Mech.*, **35,** 609–620 (1969).
117. R. C. Somerville, *J. Geophys. Fluid Dyn.*, **2,** 247–262 (1971).
118. Y. Nakagawa and P. Frenzen, *Tellus*, **7,** 1–21 (1955).
119. E. L. Koschmieder, *Beitr. Phys. Atmos.*, **40,** 216–225 (1967).
120. G. M. Homsy and J. L. Hudson, *J. Fluid Mech.*, **48,** 605–624 (1971).
121. D. J. Tritton and M. N. Zarraga, *J. Fluid Mech.*, **30,** 21–31 (1967).
122. E. W. Schwiderski and H. J. Schwab, *J. Fluid Mech.*, **48,** 703–719 (1971).
123. F. A. Kulacki and R. J. Goldstein, *J. Fluid Mech.*, **55,** 271–287 (1972).
124. P. M. Roberts, *J. Fluid Mech.*, **30,** 33–49 (1967).
125. R. Thirlby, *J. Fluid Mech.*, **44,** 673–693 (1971).
126. E. W. Schwiderski, *Phys. Fluids*, **15,** 1189–1196 (1972).
127. E. W. Schwiderski, *Phys. Fluids*, **15,** 1882–1897 (1972).
128. G. Ahlers, *Bull. Am. Phys. Soc.*, **17,** 59–60 (1972).

AN OVERVIEW OF COMPUTATIONAL METHODS FOR LARGE MOLECULES

JUERGEN HINZE

Department of Chemistry, University of Chicago
Chicago, Illinois

CONTENTS

I. Introduction 213
II. CI and SCF Formalism 217
A. Closed Shell Hartree Fock 224
B. Single Determinant Open Shell Hartree Fock 226
C. Symmetry Considerations 229
D. General Open Shell SCF Equations Including Symmetry . . . 232
E. Solving the SCF Equations 234
F. Determination of the Initial Density Matrix 237
G. Bibliography to the CI and SCF Methods 239
III. Pseudopotentials; Frozen Core and Frozen Group Approximation, Model Potentials 240
IV. Basis Sets 243
A. Type of Function 244
B. Basis Function Characteristics 246
V. Discussion of Approximate Molecular Wave Functions 247
VI. Limitation of ab initio Methods 252
VII. Semiempirical Methods 255
Acknowledgment 259
References 259

I. INTRODUCTION

The fundamental, microscopic understanding of chemistry, its processes and reactions, requires a detailed, if possible, quantitative knowledge of the properties of atoms, molecules, and their interactions. Atomic and molecular properties of chemical interest are governed by the details of their electronic structure and their nuclear arrangements. The energetics of a rearrangement of the nuclear positions—chemical reaction—are determined by the much faster rearrangement of the electronic charge cloud.

There are two important methods that permit us to obtain the necessary detailed information about the electronic structure and nuclear conformation of molecules. Spectroscopy with its many variants has yielded this

information since its inception in the discovery of absorption (emission) lines by Fraunhofer in 1815. In fact it was spectroscopy that contributed largely to the development of quantum mechanics for the interpretation of the observed effects. In the meantime quantum mechanics has matured and, with the advent of large-scale calculating machines, computational methods based on quantum mechanics have been developed which permit a detailed and quantitative description of the electronic and nuclear structure of molecules. In a sense, the evolution of theoretical methods for the elucidation of molecular structure closely parallels the earlier development of spectroscopy, and today many of the advances in one field depend on, or result in, advances of the other.

It is instructive to pursue this parallelism and the analogies following from it. Atomic spectra and the electronic structures of atoms were the earliest to be understood quantitatively. This was followed by the detailed interpretation of the band spectra of diatomics, which for the resolution of their rotation-vibrational structure required spectrometers of higher and higher resolution. Molecular orbital theory was developed for the interpretation of the spectra of diatomics, leaning heavily on the known information about the electronic structure of atoms. It has since evolved, with the help of computers and increasingly sophisticated computational methods, to a resolution closely approaching that of spectroscopy for diatomics. To be sure spectroscopy is still way ahead, but there are certain properties of interest that are now more readily attainable from calculations rather than from spectroscopy, that is, long-range forces, repulsive potential curves, electric moments of other than ground states, etc. One can expect that this list will increase and that reliable computational results will become more readily attainable with improved computational methods and computers. Both are developing rapidly.

As we proceed to larger molecules the problem of attaining detailed molecular structure information from spectroscopy and from quantum mechanical methods becomes increasingly difficult. A high-resolution UV spectrum of these molecules has so many lines that their assignment is next to impossible, and frequently the lines cannot be resolved even with the best spectrometers. But even when the Herculean task of sorting out all the lines is completed, inverting this information to obtain an energy surface, the analogue of a diatomic potential curve, is still not possible. The same holds true for many other properties of interest of polyatomic molecules. To be sure, spectroscopic methods can reveal much important electronic and nuclear structure information for these systems, although the information is not as detailed and complete as that obtainable for diatomics. However, special methods such as low-resolution UV and IR spectroscopy and particularly ESR and NMR spectroscopy are

extremely useful tools in the study of larger molecules. The situation is quite analogous in the quantum mechanical study of molecular structure. Ab initio methods, successful in atoms and small molecules, become more and more cumbersome to apply to larger systems, and one has to be satisfied with significantly lower resolution. Again it becomes advantageous to resort to special methods, like the semiempirical approaches, or model theories like pseudopotential or molecular fragment methods, which focus attention only on that part of the electronic structure that is changing in the effect studied. Of course, it is possible to carry through a full ab initio calculation of quite large molecules; however, the computational labor becomes enormous, frequently not warranted unless truly high "resolution" information is essential.

Before we go into the specific discussion of computational methods for the description of the electronic structure of large molecules, it is necessary to define what is meant here by large. To do this we classify molecules in three categories such that the classification roughly parallels the types of computational methods presently suitable for the description of the electronic structure of these molecules. This classification is best carried out in terms of N, the number of nuclei, and n, the number of electrons.

1. Small molecules and atoms, $N \leqslant 3$, $n \leqslant 20$, and $n \times N \leqslant 50$. Such systems can and should be treated with fully ab initio methods, including electron correlation. Almost any desired accuracy is attainable without excessive computational effort.

2. Medium molecules, $3 < N \leqslant 5$, $20 < n \leqslant 50$, and $50 < n \times N \leqslant 200$. Such systems can still be treated by ab initio methods restricted generally, however, to the independent particle model—the Hartree-Fock approximation—using only moderate or minimal basis sets.

3. Large molecules, $5 < N \leqslant 100$, $50 < n \leqslant 1000$, and $200 < n \times N \leqslant 10^4$. For such systems it becomes necessary to resort to semiempirical methods or model calculations of decreasing sophistication as the systems get larger.

One restriction should be mentioned here. Systems containing nuclei with nuclear charge Z larger than 50 cannot be treated to date with any reliability. The K and L shells of these atoms will be modified significantly by relativistic retardation effects and how this influences the outer valence shell is not clear yet. In general quantitative calculations of relativistic effects, including spin-orbit and spin-spin interactions, is not well developed as yet.

Clearly the limits stated above are by no means fixed; in fact one can expect with the increasing sophistication of computer programs and the

rapid advance of computer technology that these limits will shift more and more to larger n and N. Unfortunately it is true, as in spectroscopy, that those researchers who develop the increasingly sophisticated computational (spectroscopic) methods frequently will not apply them to the most interesting chemical problems or systems; that is not where their interest lies. The interest of these researchers is in developing and improving the methods, that is, designing and building better computer programs (spectrometers). Unfortunately many researchers interested in solving complex chemical problems, where electronic structure information may be important or essential, frequently use methods (computer programs or spectrometers) of much lower resolution and efficiency, because they are more readily available than the best, newest developed, and refined methods. The computational methods of quantum mechanics have definitely matured and are refined enough to yield reliable and useful molecular structure information not only for atoms and small molecules, but also for medium and large molecules. This is particularly true for the best currently developed methods, and making these methods (computer programs) available to the general research community should be an important task. To be sure, there is still much work to be done to develop and refine these methods further and to design and test new approaches in order to increase the reliability, resolution, and efficiency with which theory can predict and describe the electronic structure of molecules quantitatively.

In the specific discussion of the various computational methods for the calculation of the electronic structure of medium and large molecules, we restrict ourselves to the nonrelativistic Born-Oppenheimer (clamped nuclei) approximation. As a standard we use configuration interaction (CI) type wave functions constructed as a linear combination of antisymmetrized product functions, Slater determinants (SD's). The SD's are constructed from orthonormal single-particle functions, the orbitals, which may or may not be determined variationally. In this standard the orbitals themselves are expressed as linear combinations of some set of basis functions. After specifying the CI and SCF formalism in detail, in order to define the notation, we discuss the multitude of successful and promising methods that can be derived directly from such an ansatz by introducing restrictions and approximations. This is followed by a discussion of special methods of interest which cannot be obtained directly from such an ansatz, but we attempt to relate these procedures to our reference standard. An attempt was made to present all important and promising methods developed to date, giving an indication of their applicability and success. Clearly such a selection is subjective. Equally subjective is the bibliography, where only representative references to the various methods

are cited, rather than a complete and comprehensive bibliography, which would be forbiddingly large.

In the next section, where the general CI and SCF formalism is presented, we do not give any specific references, since it is difficult to attribute specific parts of this presentation to original papers. We attempt rather, to give a complete, comprehensive development, followed by a selected and annotated bibliography for those readers who would like to see a more detailed presentation of some of the various parts of the formalism presented.

II. CI AND SCF FORMALISM

The nonrelativistic Born-Oppenheimer Hamiltonian of an N atom, n electron molecule is

$$\mathscr{H} = \sum_{i}^{n} h(i) + \sum_{i>j}^{n} g(i,j) \tag{1}$$

with*

$$g(i,j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2}$$

the electron electron repulsion and

$$h(i) = -\tfrac{1}{2}\nabla_i^2 - \sum_{A}^{N} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} \tag{3}$$

the kinetic energy operator for electron i and the potential due to all the nuclei with nuclear charge Z_A. The total electronic energy is obtained as the eigenvalue of the Schrödinger equation for the determination of the electronic wave function

$$\mathscr{H}\Psi = E_e\Psi \tag{4}$$

In this form the wave functions Ψ and the electronic energy depend parametrically on the nuclear coordinates $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N$. The nuclear-nuclear repulsion energy

$$E_N = \sum_{A>B}^{N} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{5}$$

is a constant independent of the wave function, dependent only on the nuclear coordinates. The total energy, electronic plus nuclear repulsion energy,

$$E_T = E_e + E_N \tag{6}$$

* Atomic units are used throughout, that is $h = 1$, $e = 1$, $m = 1$, and $a_0 = 1$. The energy unit is hartree, 1 hartree $\simeq$ 27.2116 eV $\simeq$ 627.114 kcal/mole $\simeq$ 219,475 cm^{-1}, and the length unit is bohr, 1 bohr $\simeq 0.529177 \times 10^{-8}$ cm.

defines the potential for the motion of the nuclei. In general we will deal with the electronic energy E_e only, and we will drop the subscript.

In the CI form the wave function for a desired stationary state K is expanded,

$$\Psi_K = \sum_I \Phi_I C_{IK} \tag{7}$$

into antisymmetrized n-electron product functions Φ_I, that is, Slater determinants (SD's),

$$\Phi_I = \frac{1}{\sqrt{n!}} \det \{\psi_{i_1}(1)\psi_{i_2}(2) \cdots \psi_{i_n}(n)\} \tag{8}$$

which are constructed from a common set of m single particle functions, that is, orthonormal spin orbitals, ψ_i, satisfying

$$\langle \psi_i \mid \psi_j \rangle = \delta_{ij} \tag{9}$$

Choosing the spin orbitals orthonormal is no restriction in the general ansatz (7) and it results in the orthonormality of the SD's

$$\langle \Phi_I \mid \Phi_J \rangle = \delta_{IJ} \tag{10}$$

provided only nonredundant SD's are constructed. This is achieved by requiring that the set of orbitals of each Φ_I defined by

$$I \equiv \{i_1, i_2, \ldots, i_n\} \tag{11}$$

is constrained to $i_1 < i_2 < \cdots < i_n$.

Variational determination of the configuration expansion coefficients C_{IK} with fixed SD's leads to the eigenvalue equation

$$\mathbf{HC} = \mathbf{CE} \tag{12}$$

which may be solved by diagonalizing the matrix $\mathbf{H}$ defined by its elements

$$H_{IJ} = \langle \Phi_I | \mathscr{H} | \Phi_J \rangle \tag{13}$$

In (12) $\mathbf{E}$ is a diagonal matrix of the electronic energies E_K which, if ordered by increasing value, are upper bounds to the true electronic energies of the states of the system in the corresponding order.

Due to the orthonormality of the SD's, $\mathbf{C}$ is a unitary (orthogonal) matrix formed from the column vectors $\mathbf{C}_K$ ensuring the orthonormality of the final states. It should be noted here that the use of complex functions, matrices, or expansion vectors yields nothing new and is not necessary as long as we are dealing with a nonrelativistic Hamiltonian. If complex functions are used it is mostly for convenience especially in treating orbital angular momenta. Note, however, that SD's constructed from complex

spin orbitals can express specific linear combinations of SD's constructed from real orbitals.

The spin orbitals ψ used in the construction of the SD's are in general spatial functions φ multiplied by either spin function α and β. We adopt the convention

$$\begin{aligned} \psi_{2i-1}(\mathbf{r}, \mathbf{s}) &= \varphi_i(\mathbf{r})\alpha(\mathbf{s}) \\ \psi_{2i}(\mathbf{r}, \mathbf{s}) &= \varphi_i(\mathbf{r})\beta(\mathbf{s}) \end{aligned} \tag{14}$$

where each spatial orbital φ_i leads to a pair of spin orbitals having identical spatial form.

At this point it appears appropriate to emphasize that many of the conventions adopted in our definition of the reference standard are nonessential to the development of a successful computational method for the electronic structure of molecules. The specific conventions defined here, however, are commonly adopted and they form the basis of a large number of the computational methods currently in use. Nevertheless, there are several successful and promising procedures, which deviate from the conventions specified here, for example, the use of general antisymmetrized product functions built from nonorthogonal orbitals rather than SD's, or the use of different spatial orbitals for different spin, etc. We will deal with those methods in turn and will relate them to our reference standard. Also the detailed discussion of symmetry will be postponed; at this point it is not essential to our definition of the standard, except for the spin convention adopted.

The spatial orbitals are usually expanded in terms of basis functions

$$\varphi_i = \sum_{p}^{t} \chi_p c_{pi} \tag{15}$$

which are linearly independent and normalized but, in general, not orthogonal. Thus the scalar product of the orbital expansion coefficient vectors is defined in the basis function space with S the overlap matrix as a metric. The orbital orthonormality of spin orbitals with the same spin function then leads to

$$\langle \varphi_i \mid \varphi_j \rangle = \mathbf{c}_i^{+}\mathbf{S}\mathbf{c}_j = \delta_{ij} \tag{16}$$

where the overlap matrix is given by the elements

$$S_{pq} = \langle \chi_p \mid \chi_q \rangle \tag{17}$$

Spin orbitals with different spin functions are orthogonal to each other due to the integration over the spin coordinate.

Several different basis function choices are commonly employed:

1. Slater-type functions (STF's)

$$\chi_p(\mathbf{r}) = R_{\zeta n}(r_k) Y_{lm}(\theta_k \varphi_k) \tag{18}$$

with

$$R_{\zeta n}(r) = \left[\frac{(2\zeta)^{2n+1}}{(2n)!}\right]^{1/2} r^{n-1} e^{-\zeta r} \tag{19}$$

where the index p on the basis function represents the collections of constants ($\zeta nlmk$) specifying the exponent ζ, the principal quantum number n, and the azimuthal and magnetic quantum numbers l and m of the normalized spherical harmonic $Y_{lm}(\theta\varphi)$. The index k signifies the center of origin for the spherical coordinate system in terms of which the basis function is expressed, in general, the location of a constitutent atom.

2. Gaussian-type functions (GTF's)

$$\chi_p(\mathbf{r}) = N_p x_k^{\,l} y_k^{\,m} z_k^{\,n} e^{-\zeta r_k^2} \tag{20}$$

where N is a normalization constant and p stands for the collection of constants ($\zeta lmnk$). Here again k signifies the origin of the coordinate system, in general an atomic nucleus, around which the function is specified in terms of the local x, y, and z coordinates and the radial distance r. With the use of the local Cartesian coordinates atomic p-type orbitals are given by x, y, and z, respectively, multiplying $e^{-\zeta r^2}$, while d orbitals are given by x^2, y^2, z^2, xy, yz, and zx multiplying $e^{-\zeta r^2}$. However, there is not a one-to-one correspondence of these functions and atomic orbitals defined in terms of spherical harmonics; for example, the six d-type functions given above contain the five independent atomic d functions and an s function. Similarly, redundant functions are obtained for orbitals of higher azimuthal quantum number, the number of which is generally given here by the sum of exponents in (20), $l + m + n$.

3. Gaussian-lobe functions (GLF's)

$$\chi_p(\mathbf{r}) = N_p e^{-\zeta r_k^2} \tag{21}$$

which are the same as the GTF's except that they have no explicit angular dependence. This is compensated for by having them localized on centers other than the positions of the atomic nuclei.

We discuss the merits of the different basis function choices as we proceed and note here only that (*a*) we will need at least as many linearly independent basis functions, t, as we have spatial orbitals, that is, $m/2$, and (*b*) many of the semiempirical methods applied to truly large molecules make only implicit reference to basis functions without requiring their

specific definition, since molecular integrals are not evaluated by explicit quadrature.

We have seen how the configuration expansion coefficients C_{IK} of (7) may be obtained variationally yielding (12), which can be solved easily by standard matrix diagonalization techniques. This variational solution requires the SD's and thus the orbitals used in their construction to remain fixed. Clearly such a configuration interaction calculation depends critically on the types of SD's selected to participate in the expansion, (7), as well as on the appropriate determination of the orbitals constructing the SD's. Only a careful determination of these two variables will result in a final total wave function that realistically describes the electronic structure of the desired molecular states in tractable form, that is, not requiring thousands or millions of terms in the expansion of (7). Again we can make use of the variational principle for the determination of the best orbitals to be used, that is, the expansion coefficients in (15), in the construction of the total wave function. The familiar result of such a variational determination of the orbitals are the Hartree-Fock (HF) equations that are obtained if the expansion, (7), of the total wave function is restricted to a single SD. We do not make such a restriction here, but rather derive the fully general multiconfiguration self-consistent field (MCSCF) orbital equations, the open and closed shell HF equations being a special case. Before we can apply the variational principle with respect to a variation of the orbitals to the expectation value of the energy

$$E_K = \langle \Psi_K | \mathscr{H} | \Psi_K \rangle = \sum_{IJ} C_{KI}{}^* \langle \Phi_I | \mathscr{H} | \Phi_J \rangle C_{KJ} \tag{22}$$

it is necessary to express E_K such that the orbital dependence appears explicitly. To this end it is convenient to make use of the general first- and second-order reduced density kernels defined as

$$\gamma(1, 1') = n \int \Psi(1, 2, \ldots, n) \Psi^*(1', 2, \ldots, n)\, d2 \cdots n \tag{23}$$

and

$$\Gamma(1, 2; 1', 2') = n(n-1) \int \Psi(1, 2, \ldots, n) \Psi^*(1', 2', \ldots, n)\, d3 \cdots n \tag{24}$$

These may be expressed in terms of the spin orbitals as

$$\gamma(1, 1') = \sum_{ij}^{m} \gamma_{ij} \psi_j(1) \psi_i{}^*(1') \tag{25}$$

and

$$\Gamma(1, 2; 1'2') = \sum_{ijkl}^{m} \Gamma_{ij,kl} \psi_j(1) \psi_l(2) \psi_i{}^*(1') \psi_k{}^*(2') \tag{26}$$

where γ_{ij} and $\Gamma_{ij,kl}$ are the first- and second-order reduced density matrices of state Ψ in the space spanned by the spin orbitals. Since we are focusing on one state only, we have dropped the state index K.

Note that the total first-order reduced density kernel may be obtained in terms of the transition density kernels between the individual SD's defined as

$$\gamma_{IJ}(1, 1') = n\int \Phi_J(1, \ldots, n)\Phi_I^*(1', \ldots, n)\, d2 \cdots n$$
$$= \sum_{ij}^{m} \gamma_{IJ,ij}\psi_j(1)\psi_i^*(1') \tag{27}$$

by summing over the SD's with the given configuration expansion coefficients

$$\gamma_{ij} = \sum_{IJ} \gamma_{IJ,ij} C_{IK}^* C_{JK} \tag{28}$$

and similarly for the second-order reduced density matrix.

Thus the energy matrix elements are in terms of the spin orbitals

$$H_{IJ} = \langle \Phi_I | \mathscr{H} | \Phi_J \rangle = \sum_{ij}^{m} \gamma_{IJ,ij} \langle \psi_i | h | \psi_j \rangle + \tfrac{1}{2} \sum_{ijkl}^{m} \Gamma_{IJ,ij,kl} \langle \psi_i \langle \psi_k | g | \psi_l \rangle \psi_j \rangle \tag{29}$$

and the expectation value of the energy is

$$E = \sum_{ij}^{m} \gamma_{ij} \langle \psi_i | h | \psi_j \rangle + \tfrac{1}{2} \sum_{ijkl}^{m} \Gamma_{ij,kl} \langle \psi_i \langle \psi_k | g | \psi_l \rangle \psi_j \rangle \tag{30}$$

Variation of this expression with respect to a spin orbital change after adding the restrictive orthonormality conditions

$$\langle \psi_i \mid \psi_j \rangle = \delta_{ij} \tag{31}$$

multiplied by Lagrangian multipliers ϵ_{ji} leads to the Fock-like MCSCF equations

$$\sum_{j}^{m} F_{ij}\psi_j = \sum_{j}^{m} \psi_j \epsilon_{ji} \tag{32}$$

for the determination of the orbitals. Here the Fock-like operators are given by

$$F_{ij} = \gamma_{ij} h + \sum_{kl}^{m} \Gamma_{ij,kl} \langle \psi_k | g | \psi_l \rangle \tag{33}$$

where the last term after integration over coordinate 2 is the potential as a function of coordinate 1.

Equations (29), (30), (32), and (33) can be expressed readily in terms of the spatial orbitals φ by defining the density matrices after integration over the spin coordinate; that is,

$$\gamma'(1, 1') = n\int \Psi(1, 2, \ldots, n)\Psi^*(1', 2, \ldots, n)\, ds_1, 2, \ldots, n$$
$$= \sum_{ij}^{m/2} \gamma_{ij}'\varphi_j(1)\varphi_i^*(1') \tag{34}$$

Similarly the second-order reduced density matrix and the transition density matrices may be defined in terms of the spatial orbitals. With these we obtain in terms of the spatial orbitals

$$\langle \Phi_I | \mathscr{H} | \Phi_J \rangle = \sum_{ij}^{m/2} \gamma'_{IJ,ij} \langle \varphi_i | h | \varphi_j \rangle + \tfrac{1}{2} \sum_{ijkl}^{m/2} \Gamma'_{IJ,ij,kl} \langle \varphi_i \langle \varphi_k | g | \varphi_l \rangle \varphi_j \rangle \tag{29a}$$

$$E = \sum_{ij}^{m/2} \gamma_{ij}' \langle \varphi_i | h | \varphi_j \rangle + \tfrac{1}{2} \sum_{ijkl}^{m/2} \Gamma'_{ij,kl} \langle \varphi_i \langle \varphi_k | g | \varphi_l \rangle \varphi_j \rangle \tag{30a}$$

This leads to the MCSCF equations

$$\sum_{j}^{m/2} F_{ij}' \varphi_j = \sum_{j}^{m/2} \varphi_j \epsilon_{ji} \tag{32a}$$

with

$$F_{ij}' = \gamma_{ij}' h + \sum_{kl}^{m/2} \Gamma'_{ij,kl} \langle \varphi_k | g | \varphi_l \rangle \tag{33a}$$

It is seen that all equations in terms of the spatial orbitals have identical form to the corresponding equations in terms of the spin orbitals except that the reduced density matrices γ' and Γ' are expressed in the space spanned by the spatial orbitals. From hereon we deal with spatial orbitals only and drop the primes.

Expressing (29a), (30a), (32a), and (33a) in terms of the basis function expansion (15) is a straightforward substitution yielding

$$\langle \Phi_I | \mathscr{H} | \Phi_J \rangle = \sum_{ij}^{m/2} \gamma_{IJ,ij} \sum_{pq}^{t} c_{pi} \langle \chi_p | h | \chi_q \rangle c_{qj}$$
$$+ \tfrac{1}{2} \sum_{ijkl}^{m/2} \Gamma_{IJ\ ij,kl} \sum_{pqrs}^{t} c_{pi} c_{rk} \langle \chi_p \langle \chi_r | g | \chi_s \rangle \chi_q \rangle c_{sl} c_{qj} \tag{29b}$$

$$E = \sum_{ij}^{m/2} \gamma_{ij} \sum_{pq}^{t} c_{pi} \langle \chi_p | h | \chi_q \rangle c_{qj}$$
$$+ \tfrac{1}{2} \sum_{ijkl}^{m/2} \Gamma_{ij,kl} \sum_{pqrs}^{t} c_{pi} c_{rk} \langle \chi_p \langle \chi_r | g | \chi_s \rangle \chi_q \rangle c_{sl} c_{qj} \tag{30b}$$

and the MCSCF equations

$$\sum_{j} \mathbf{F}_{ij} \mathbf{c}_j = \mathbf{S} \sum_{j} \mathbf{c}_j \epsilon_{ji} \tag{32b}$$

with

$$F_{ij,pq} = \gamma_{ij}\langle\chi_p| h |\chi_q\rangle + \sum_{kl}^{m/2} \Gamma_{ij,kl} \sum_{rs}^{t} c_{rk}\langle\chi_p\langle\chi_r| g |\chi_s\rangle\chi_q\rangle c_{sl} \tag{33b}$$

At this point the reader may wonder why we have presented here the most general equations, which permit the calculation of correlated electronic wave functions using conventional CI techniques or the MCSCF method, rather than just giving the equations for the Hartree-Fock type wave functions most commonly used in calculations of large molecules. This was done for several reasons: (*a*) the equations are really not much more complicated. (*b*) It has been found in many calculations on small molecules that the explicit inclusion of correlation effects, at least in the valence shell, is essential for the quantitative ab initio calculation of dissociation and electronic excitation energies and other molecular properties. (*c*) It can therefore be expected that electronic wave function calculations for medium and larger molecules will make increasing use of these methods. (*d*) The expressions for the matrix elements $\langle\Phi_I| \mathscr{H} |\Phi_J\rangle$ are generally needed in the description of excited states using the CI formalism. (*e*) Finally it is relatively easy to make the transition from these general equations to the equations of the conventional restricted closed or open shell Hartree Fock model commonly used.

A. Closed Shell Hartree Fock

In the Hartree-Fock (independent particle) model the wave function is approximated by a single configuration. In general it will be necessary for symmetry reasons to express such a single configuration as a specific linear combination of SD's. For the time being we restrict ourselves to the special case where the configuration can be expressed by a single SD. We deal with the general case when we consider symmetry explicitly. First we consider the simplest, most common closed shell case where we have n = even electrons occupying pairwise the α and β spin-orbitals formed from $n/2$ spatial orbitals, giving a singlet state. In this case the only non-zero first- and second-order reduced density matrix elements are for i and j less than or equal to $n/2$

$$\gamma_{ii} = 2$$

$$\Gamma_{ii,jj} = 4 \tag{35}$$

and

$$\Gamma_{ij,ji} = -2$$

Furthermore, it is possible to use the freedom of a unitary transformation among the occupied $n/2$ spatial orbitals to eliminate the off-diagonal Lagrangian multipliers. Using this in the general equations above we

obtain

$$E = \langle \Phi_1 | \mathscr{H} | \Phi_1 \rangle = \sum_i^{n/2} 2\langle \varphi_i | h | \varphi_i \rangle + \tfrac{1}{2} \sum_{ij}^{n/2} (4\langle \varphi_i \langle \varphi_j | g | \varphi_j \rangle \varphi_i \rangle - 2\langle \varphi_i \langle \varphi_j | g | \varphi_i \rangle \varphi_j \rangle) \quad (36)$$

and the Hartree-Fock equations

$$F\varphi_i = \epsilon_{ii}\varphi_i \quad (37)$$

with

$$F = 2h + \sum_j^{n/2} (4\langle \varphi_j | g | \varphi_j \rangle - 2\langle \varphi_j | gP_{12} | \varphi_j \rangle) \quad (38)$$

Note that the Fock operator of (38) is twice the Fock operator of the conventional definition. This is to keep it equivalent to the general definition of (32) and (33) which has distinct advantages in the open shell case. Normalizing the Fock operator to the shell occupation number, γ_{ii}, rather than to one, will not affect the orbitals; however, our orbital energies ϵ_{ii} will be γ_{ii} times the conventional orbital energies. Furthermore we have introduced the exchange operator P_{12} which exchanges particle coordinates 1 and 2 before the integration can be performed. This "trick" permits the writing of a single Fock operator in (37) rather than the sum over j as we had in (32). With this exchange operator P_{12} we have the customary nonlocal exchange potential defined as

$$K_j | \varphi \rangle = \langle \varphi_j | gP_{12} | \varphi_j \rangle | \varphi \rangle = \langle \varphi_j | g | \varphi \rangle | \varphi_j \rangle \quad (39)$$

Using the basis function expansion of the spatial orbitals, (15), we obtain

$$E = \sum_{pq}^{t} D_{pq}\langle \chi_p | h | \chi_q \rangle + \tfrac{1}{2} \sum_{pq,rs}^{t} D_{pq}D_{rs}(\langle \chi_p \langle \chi_r | g | \chi_s \rangle \chi_q \rangle - \tfrac{1}{2}\langle \chi_p \langle \chi_r | g | \chi_q \rangle \chi_s \rangle) \quad (36a)$$

and

$$\mathbf{Fc} = \mathbf{Sc}\boldsymbol{\epsilon} \quad (37a)$$

where $\mathbf{c}$ is the matrix formed by the column vectors, $\mathbf{c}_i$ of the orbital expansion coefficients of orbital i, and $\boldsymbol{\epsilon}$ is the diagonal matrix of the orbital energies. The Fock matrix $\mathbf{F}$ is obtained as

$$F_{pq} = 2\langle \chi_p | h | \chi_q \rangle + \sum_{rs}^{t} D_{rs}(2\langle \chi_p \langle \chi_r | g | \chi_s \rangle \chi_q \rangle - \langle \chi_p \langle \chi_r | g | \chi_q \rangle \chi_s \rangle) \quad (38a)$$

In the equations above it was convenient to use the first-order reduced density matrix in the space spanned by the basis functions which is given as

$$D_{pq} = \sum_i^{n/2} 2c_{pi}c_{qi} \quad (40)$$

B. Single Determinant Open Shell Hartree Fock

To obtain the HF equations for a nonclosed shell, nonsinglet state, we adopt the convention that the first k orbitals are doubly occupied and the remaining electrons are described by $l = n - 2k$ spin orbitals all with α spin. A single determinant built from these orbitals will be an eigenfunction of S^2 with $S = M_S = l/2$. In this case we obtain the following nonzero values for the first- and second-order reduced density matrices:

$$\gamma_{ii} = 2 \qquad \text{for } i \leqslant k$$
$$\gamma_{ii} = 1 \qquad \text{for } k < i \leqslant k + l \tag{41}$$
$$\Gamma_{ii,jj} = 4 \quad \text{and} \quad \Gamma_{ij,ji} = -2 \quad \text{for } i \text{ and } j \leqslant k$$
$$\Gamma_{ii,jj} = 2 \quad \text{and} \quad \Gamma_{ij,ji} = -1 \quad \text{for } i \leqslant k \text{ and } k < j \leqslant l$$
$$\text{or } k < i \leqslant l \text{ and } k < j \leqslant l$$
$$\Gamma_{ii,jj} = 1 \quad \text{and} \quad \Gamma_{ij,ji} = -1 \quad \text{for } k < i \leqslant l \text{ and } k < j \leqslant l$$

We obtain for the energy expression

$$E = \sum_{i}^{k+l} \gamma_{ii} \langle \varphi_i | h | \varphi_i \rangle + \tfrac{1}{2} \sum_{ij}^{k+l} (\Gamma_{ii,jj} \langle \varphi_i \langle \varphi_j | g | \varphi_j \rangle \varphi_i \rangle + \Gamma_{ij,ji} \langle \varphi_i \langle \varphi_j | g | \varphi_i \rangle \varphi_j \rangle) \tag{42}$$

In the derivation of the Fock equations for this case, that is, substituting the density matrices into the general (32), we are left with two equations, one for the doubly occupied orbitals φ_1 through φ_k and another for the singly occupied orbitals φ_{k+1} through φ_{k+l}. Furthermore, it is now not possible to eliminate all the off-diagonal Lagrangian multipliers, since only unitary transformations among the doubly occupied orbitals and among the single occupied orbitals leave the wave function invariant, but not a unitary transformation between doubly and singly occupied orbitals. Thus we are left with off-diagonal Lagrangian multipliers between doubly and singly occupied orbitals. The Fock equations for the doubly occupied orbitals are

$$F_{\mathscr{C}} \varphi_i = \varphi_i \epsilon_{ii} + \sum_{j=k+1}^{k+l} \varphi_j \epsilon_{ji} \qquad i \leqslant k \tag{43}$$

and for the singly occupied orbitals

$$F_{\mathscr{S}} \varphi_i = \varphi_i \epsilon_{ii} + \sum_{j=1}^{k} \varphi_j \epsilon_{ji} \qquad k < i \leqslant l \tag{44}$$

with the Fock operator

$$F_{\mathscr{C}} = 2h + \sum_{j=1}^{k}(4\langle\varphi_j|\, g\,|\varphi_j\rangle - 2\langle\varphi_j|\, gP_{12}\,|\varphi_j\rangle) + \sum_{j=k+1}^{k+l}(2\langle\varphi_j|\, gP_{12}\,|\varphi_j\rangle) \tag{45}$$

and

$$F_{\mathscr{S}} = h + \sum_{j=1}^{k}(2\langle\varphi_j|\, g\,|\varphi_j\rangle - \langle\varphi_j|\, gP_{12}\,|\varphi_j\rangle) + \sum_{j=k+1}^{k+l}(\langle\varphi_j|\, g\,|\varphi_j\rangle - \langle\varphi_j|\, gP_{12}\,|\varphi_j\rangle) \tag{46}$$

We can now realize the advantage of normalizing the Fock operators to γ_{ii} rather than one, since with our normalization we have $\epsilon_{ij} = \epsilon_{ji}$ if (43) and (44) are satisfied.

In order to express (42) through (46) in terms of basis functions it is convenient to define three density matrices:

1. For the completely, doubly occupied shells

$$D_{\mathscr{C},pq} = \sum_{i=1}^{k} 2c_{pi}c_{qi} \tag{47}$$

2. For the singly occupied orbitals

$$D_{\mathscr{S},pq} = \sum_{i=k+1}^{k+l} c_{pi}c_{qi} \tag{48}$$

3. For the total wave function

$$D_{\mathscr{T},pq} = D_{\mathscr{C},pq} + D_{\mathscr{S},pq} \tag{49}$$

With these definitions we obtain for the total energy

$$E = \sum_{pq} D_{\mathscr{T},pq}\langle\chi_p|\, h\,|\chi_q\rangle + \tfrac{1}{2}\sum_{pqrs}[D_{\mathscr{T},pq}D_{\mathscr{T},rs}(\langle\chi_p\langle\chi_r|\, g\,|\chi_s\rangle\chi_q\rangle - \tfrac{1}{2}\langle\chi_p\langle\chi_r|\, g\,|\chi_q\rangle\chi_s\rangle) - \tfrac{1}{2}D_{\mathscr{S},pq}D_{\mathscr{S},rs}\langle\chi_p\langle\chi_r|\, g\,|\chi_q\rangle\chi_s\rangle] \tag{42a}$$

and for the Fock equations

$$\mathbf{F}_{\mathscr{C}}c_i = \mathbf{S}\left(\mathbf{c}_i\epsilon_{ii} + \sum_{j=k+1}^{l+1}\mathbf{c}_j\epsilon_{ji}\right) \qquad \text{for } i \leqslant k \tag{43a}$$

and

$$\mathbf{F}_{\mathscr{S}}c_i = \mathbf{S}\left(\mathbf{c}_i\epsilon_{ii} + \sum_{j=1}^{k}c_j\epsilon_{ji}\right) \qquad \text{for } k < i \leqslant l \tag{44a}$$

with the definition of the Fock operators

$$F_{\mathscr{C},pq} = 2\langle\chi_p| h |\chi_q\rangle + \sum_{rs} D_{\mathscr{T},rs}(2\langle\chi_p\langle\chi_r| g |\chi_s\rangle\chi_q\rangle - \langle\chi_p\langle\chi_r| g |\chi_q\rangle\chi_s\rangle) \tag{45a}$$

and

$$F_{\mathscr{S},pq} = \langle\chi_p| h |\chi_q\rangle + \sum_{rs} [D_{\mathscr{T},rs}\langle\chi_p\langle\chi_r| g |\chi_s\rangle\chi_q\rangle - \tfrac{1}{2}(D_{\mathscr{T},rs} + D_{\mathscr{S},rs})\langle\chi_p\langle\chi_r| g |\chi_q\rangle\chi_s\rangle] \tag{46a}$$

It is not difficult to solve the closed shell HF equations, (37a). This is accomplished by diagonalizing the Fock matrix, (38a). The presence of the overlap matrix in (37a) presents little extra difficulty. Clearly the equations have to be solved iteratively since the Fock operator depends on the form of the occupied orbitals. Even that presents, in general, no problem. The same is not true for the solution of the open shell HF equations, (43) and (44). The nonzero off-diagonal Lagrangian multipliers present some extra problems; solution of these equations by direct diagonalization of the Fock matrices is not possible. Before we discuss these matters in detail, however, we should realize that the open shell system dealt with explicitly is of a very special nature and by no means general, since we have required that the singly occupied orbitals used in the construction of our single configuration function all have the same spin part. It was this restriction that allowed us to transform away the off-diagonal Lagrangian multipliers not only between closed shells but also between open shells, leaving only those between open and closed shells. Frequently, in particular in the description of excited states, it is necessary to allow for more general single configuration functions which, in general, cannot be expressed in terms of a single SD. It is true that the general SCF equations given, (32) and (33), permit us to deal with all these cases; however, their solution is by no means trivial. There are still significant simplifications obtained if we deal with a single configuration HF type problem, even though such a configuration may require several SD's. Since such a situation is generally closely connected with the request that out wave function describe a state of specific symmetry, it is necessary here to deal explicitly with such symmetry requirements. In fact we have made use of such symmetry requirements already. Requesting that the same spatial orbital be multiplied by either spin function and used in pairs in the closed shell case was just an easy way to ensure that our single SD's would be eigenfunctions of S^2.

C. Symmetry Considerations

The symmetry operators of the point group of the nuclear conformation of a given molecule and the spin operators S^2 and S_z commute with the Hamiltonian defined in (1). Therefore, eigenfunctions of $\mathscr{H}$ will also be eigenfunctions of these symmetry operators, the corresponding eigenvalues being constants of motion. These are generally used as quantum numbers characterizing a specific stationary state. In the case of spin we have the value of S or $(2S + 1)$, the multiplicity of the state, and its M_S component. In the case of the collection of symmetry operators of the point group of the molecule we generally choose the irreducible representations of the group. States with different sets of such quantum numbers will not interact; that is, if the set of these quantum numbers is denoted by Λ then

$$\langle \Psi_\Lambda | \mathscr{H} | \Psi_{\Lambda'} \rangle = 0 \qquad \text{if } \Lambda \neq \Lambda' \tag{50}$$

Since the electronic states of molecules are signified by such sets of quantum numbers, and since the use of such noninteracting functions in general reduces the dimensionality of the computational problems, it is convenient and customary to make use of symmetry in the construction of wave functions. In addition to spin we have made use already of another symmetry, that is, permutational symmetry. By choosing SD's for the configuration expansion of the wave functions (7), we have ensured that our wave functions are antisymmetric with respect to electron interchange; that is, they transform as the totally antisymmetric irreducible representation of the permutation group of n particles S_n, as required by the Pauli principle. To ensure that our total wave functions transform as irreducible representations of the molecular point group and are eigenfunctions of S^2 and M_S, we place these requirements on each individual configuration used to expand the wave functions. In general, it is not possible to satisfy these requirements by using only a single SD to express a configuration. However, it is always possible to find a specific linear combination of SD's satisfying the symmetry requirements. Such functions we call configuration state functions (CSF's).

The CSF's are most easily and conventionally constructed by choosing symmetry adapted orbitals as their building blocks. Thus we require that our spatial orbitals transform as irreducible representations of the point group of the system. Signifying their symmetry species by $\lambda, \mu, \ldots$ and subspecies by $\alpha, \beta, \ldots$, the basis function expansion, (15), for the symmetry adapted orbitals becomes

$$\varphi_{\lambda\alpha i} = \sum_p \chi_{\lambda\alpha p} c_{\lambda p i} \tag{51}$$

where we have used symmetry adapted basis functions in the expansion.

In general, these symmetry adapted basis functions transform as irreducible representation λ with subspecies α of the point group of the system. They will be specific linear combinations of the primitive basis functions introduced above, (18) through (21). In (51) the indices i and p differentiate orbitals and basis functions of identical symmetry, respectively, and they range over the dimensionality of the orbitals m_λ and basis functions t_λ within that symmetry λ. In the case of degenerate representations λ of order g_λ, the index of the subspecies, α, ranges over this dimensionality. For our CSF's, we require that the basis functions $\chi_{\lambda\alpha p}$ for $\alpha = 1, \ldots, g_\lambda$ with fixed λ and p span the full degenerate representation λ; that is, they are essentially equivalent. With this restriction the orbitals $\varphi_{\lambda\alpha i}$ for $\alpha = 1, \ldots, g_\lambda$ also span the full degenerate representation λ since the expansion coefficient $c_{\lambda pi}$ have been chosen independent of the subspecies α. It may be worth noting here that we have used such an equivalence restriction already by requiring that our spin orbitals, symmetry species $S = \frac{1}{2}$ and subspecies $M_S = \alpha = \frac{1}{2}$ and $M_S = \beta = -\frac{1}{2}$, be constructed pairwise with the same spatial function.

A configuration is conventionally defined in terms of such symmetry and equivalence restricted orbitals by specifying the *shell* occupation numbers $n_{\lambda i}$, that is $1s^2 2s^2 2p^2$ for the ground configuration of carbon, $1\sigma^2 2\sigma^2 3\sigma^2 1\pi$ for CH, or $1a_1^2 2a_1^2 1e^4 3a_1^2$ for NH_3. The shells are specified by $(i\lambda)$, with i preceding the symmetry specification given in terms of the symbols used to identify the irreducible representation of the point group, while the shell occupation numbers $n_{\lambda i}$ are given as n superscript.

In general it is possible to generate several distinct CSF's from a given configuration defined in terms of the symmetry and equivalence restricted orbitals introduced above, for example, the different term values arising from a given atomic configuration. We do not elaborate on the details of constructing the CSF's; this has been presented explicitly for atoms in the books by Condon and Shortley and by Slater. The application of these conventional methods to molecules, which have lower symmetry, is quite straightforward. Rather we discuss the particular form obtained for the energy expressions and the Fock operator if symmetry of the orbitals and basis functions is considered explicitly. Here is where the real computational saving lies, since the number of integrals to be computed and the size of the matrices to be handled are reduced drastically by the explicit use of symmetry.

All one-electron operators needed in our energy calculation are totally symmetric; therefore matrix elements connecting basis functions of different symmetry will be zero, and we need only the following integrals over the basis functions:

$$S_{\lambda pq} = \langle \chi_{\lambda p} \mid \chi_{\lambda q} \rangle \tag{52}$$

and

$$H_{\lambda p q} = \langle \chi_{\lambda p} | \, h \, | \chi_{\lambda q} \rangle \tag{53}$$

These integrals are independent of the choice of basis function subspecies due to the equivalence restriction; that is,

$$H_{\lambda p q} = \langle \chi_{\lambda \alpha p} | \, h \, | \chi_{\lambda \alpha q} \rangle = \langle \chi_{\lambda \beta p} | \, h \, | \chi_{\lambda \beta q} \rangle, \text{ etc.} \tag{54}$$

while

$$\langle \chi_{\lambda \alpha p} | \, h \, | \chi_{\lambda \beta q} \rangle = 0 \qquad \text{if } \alpha \neq \beta \tag{55}$$

The situation is somewhat more complicated for two-electron integrals due to the possibility of degenerate representations. Again the operator $g(ij)$ is totally symmetric and the two-electron integral $\langle \chi_{\lambda \alpha p} \langle \chi_{\rho \gamma r} | \, g \, | \chi_{\sigma \delta s} \rangle \chi_{\mu \beta q} \rangle$ can be considered as an integral of the potential $\langle \chi_{\rho \gamma r} | \, g \, | \chi_{\sigma \delta s} \rangle$ with the charge distribution $\chi^*_{\lambda \alpha p} \chi_{\mu \beta q}$. Both can be expanded into linear combinations of irreducible representations of species ν and subspecies η; that is,

$$\langle \chi_{\rho \gamma r} | \, g \, | \chi_{\sigma \delta s} \rangle = \sum_{\nu \eta} C(\rho \gamma, \sigma \delta; \nu \eta) \Omega_{\nu \eta, \rho r \sigma s} \tag{56}$$

and

$$\chi^*_{\lambda \alpha p} \chi_{\mu \beta q} = \sum_{\nu' \eta'} C^*(\lambda \alpha, \mu \beta; \nu' \eta') \omega^*_{\nu' \eta', \lambda p \mu q} \tag{57}$$

The only nonzero, nonredundant two-electron integrals are obtained when $\nu = \nu'$ and $\eta = \eta'$, and these integrals are independent of η, giving

$$R_{\nu, \lambda p \mu q, \rho r \sigma s} = \langle \omega_{\nu \eta, \lambda p \mu q} \, | \, \Omega_{\nu \eta, \rho r \sigma s} \rangle \tag{58}$$

In short, we have nonzero two-electron integrals as given in (58) only if the irreducible representation ν is contained in both the direct products $\lambda \times \mu$ and $\rho \times \sigma$.

With the definition of first- and second-order reduced density matrices in the space spanned by the symmetry orbitals, analogously to (34), as $\gamma_{\lambda i j}$ and $\Gamma_{\nu, \lambda i \mu j, \rho k \sigma l'}$ we are ready to write out explicitly the energy expressions and Fock equations for symmetry orbitals. It should be noted here that these density matrix elements are again independent of the subspecies. Furthermore we have dealt here with the complications arising from the presence of degenerate representations in a rather cursory manner; a necessary more detailed treatment would lead us too far afield.

Using symmetry explicitly together with the definitions above, we obtain for the general energy matrix elements between CSF's

$$\langle \Phi_I | \, \mathscr{H} \, | \Phi_J \rangle = \sum_{\lambda i j} \gamma_{IJ, \lambda i j} H_{\lambda i j} + \tfrac{1}{2} \sum_{\nu, \lambda i \mu j, \rho k \sigma l} \Gamma_{IJ, \nu, \lambda i \mu j, \rho k \sigma l} R_{\nu, \lambda i \mu j, \rho k \sigma l} \tag{59}$$

where $H_{\lambda i j}$ and $R_{\nu, \lambda i \mu j, \rho k \sigma l}$ are defined as in (54) and (58) in terms of the orbitals rather than the basis functions. For the expectation value of the

total electronic energy we have

$$E = \sum_{IJ} C_I^* C_J \langle \Phi_I | \mathscr{H} | \Phi_J \rangle$$

$$= \sum_{\lambda ij} \gamma_{\lambda ij} H_{\lambda ij} + \tfrac{1}{2} \sum_{\nu,\lambda i\mu j,\rho k\sigma l} \Gamma_{\nu,\lambda i\mu j,\rho k\sigma l} R_{\nu,\lambda i\mu j,\rho k\sigma l} \tag{60}$$

Expressed in terms of the orbital expansion coefficients and basis function integrals this becomes

$$E = \sum_{\lambda pq} D_{\lambda pq} H_{\lambda pq} + \tfrac{1}{2} \sum_{\nu,\lambda p\mu q,\rho r\sigma s} D_{\nu,\lambda p\mu q,\rho r\sigma s} R_{\nu,\lambda p\mu q,\rho r\sigma s} \tag{61}$$

where we have used the first- and second-order reduced density matrices in the space spanned by the basis functions defined respectively as

$$D_{\lambda pq} = \sum_{i,j}^{n_\lambda} \gamma_{\lambda ij} c_{\lambda pi} c_{\lambda qj} \tag{62}$$

and

$$D_{\nu,\lambda p\mu q,\rho r\sigma s} = \sum_{i}^{n_\lambda} \sum_{j}^{n_\mu} \sum_{k}^{n_\rho} \sum_{l}^{n_\sigma} \Gamma_{\nu,\lambda i\mu j,\rho k\sigma l} c_{\lambda pi} c_{\mu qj} c_{\rho rk} c_{\sigma sl} \tag{63}$$

D. General Open Shell SCF Equations Including Symmetry

It is now straightforward to write down the general MCSCF equations, [see (32) and (33)] for the determination of the symmetry adapted orbitals of a wave function built as a linear combination of the CSF's. We do not do this, but rather discuss the simplification obtained if the wave function is approximated by a single CSF, obtaining the equations for the symmetry orbitals in the general open shell HF case. Before we do this it is convenient to consider separately closed and open shell orbitals.

We define as closed shell those orbitals that describe a shell fully occupied, that is, occupied with $d_\lambda = 2g_\lambda$ electrons, with g_λ the degeneracy of the representation λ. The remaining orbitals occupied in the CSF are called open shell orbitals. It is convenient furthermore to define an electron interaction matrix containing coulomb and exchange parts as

$$P_{\lambda pq,\rho rs} = R_{0,\lambda p\lambda q\rho r\rho s} + \sum_{\nu} x_{\nu\lambda\rho} R_{\nu,\lambda p\rho s,\rho r\lambda q} \tag{64}$$

where $\nu = 0$ is the totally symmetric representation and the $x_{\nu\lambda\rho}$'s are simple numbers depending on the shell occupation numbers of λ and ρ only, defined such that the electron interaction between closed shells and between closed and open shells can be described completely by P.

The expectation value for the total energy is given in the case of a single CSF as

$$E = E_{\mathscr{C}} + \sum_{\lambda i \subset \mathscr{O}} \gamma_{\lambda ii} G_{\lambda ii} + \tfrac{1}{2} \sum_{\lambda i \rho k \subset \mathscr{O}} \Gamma_{0,\lambda i \lambda i, \rho k \rho k} R_{0,\lambda i \lambda i, \rho k \rho k} + \tfrac{1}{2} \sum_{\nu, \lambda i \rho k \subset \mathscr{O}} \Gamma_{\nu,\lambda i \rho k, \rho k \lambda i} R_{\nu,\lambda i \rho k, \rho k \lambda i} \quad (65)$$

where the explicit sums are over the valence or open shells only, with the last two terms giving the detailed interaction of the open shell electrons with each other. The matrix element $G_{\lambda ii}$ gives the energy of an open shell electron in the field of the nuclei and the core electrons; it is obtained as

$$G_{\lambda ii} = H_{\lambda ii} + \sum_{\rho k \subset \mathscr{C}} d_\rho P_{\lambda ii, \rho kk} \quad (66)$$

where d_ρ is the shell occupation number of the closed shell ρ. Finally the energy of the closed shell is given as

$$E_{\mathscr{C}} = \sum_{\lambda i \subset \mathscr{C}} d_\lambda \left(H_{\lambda ii} + \tfrac{1}{2} \sum_{\rho k \subset \mathscr{C}} d_\rho P_{\lambda ii, \rho kk} \right) \quad (67)$$

which is the only part needed for a closed shell state. It should be noted here that the energy expression, (65), also holds for a many configurational wave function, where the CSF's correspond to configurations all differing from each other by even shell occupation numbers.

The Fock equations corresponding to the energy expression, (65), are for the closed shell orbitals

$$F_{\mathscr{C},\lambda} \varphi_{\lambda i} = \varphi_{\lambda i} \epsilon_{\lambda ii} + \sum_{j \subset \mathscr{O}_\lambda} \varphi_{\lambda i} \epsilon_{\lambda ji} \quad (68)$$

where the sum on the right is over the open shell orbitals of symmetry λ. For the open shell orbitals we obtain

$$F_{\mathscr{O},\lambda i} \varphi_{\lambda i} = \sum_{j \subset \lambda} \varphi_{\lambda j} \epsilon_{\lambda ji} \quad (69)$$

In basis function expansion form these Fock equations become

$$\mathbf{F}_{\mathscr{C},\lambda} \mathbf{c}_{\lambda i} = \mathbf{S}_\lambda \left(\mathbf{c}_{\lambda i} \epsilon_{\lambda ii} + \sum_{j \subset \mathscr{O}_\lambda} \mathbf{c}_{\lambda j} \epsilon_{\lambda ji} \right) \quad (68a)$$

and

$$\mathbf{F}_{\mathscr{O},\lambda i} \mathbf{c}_{\lambda i} = \mathbf{S}_\lambda \sum_{j \subset \lambda} \mathbf{c}_{\lambda j} \epsilon_{\lambda ji} \quad (69a)$$

with

$$F_{\mathscr{C},\lambda pq} = d_\lambda \left(H_{\lambda pq} + \sum_{\rho rs} D_{\mathscr{T},\rho rs} P_{\lambda pq, \rho rs} \right) \quad (70)$$

and

$$F_{\mathcal{O},\lambda ipq} = \gamma_{\lambda ii} G_{\lambda pq} + \sum_{\rho k \subset \mathcal{O}} \sum_{rs} \Big(\Gamma_{0,\lambda i\lambda i,\rho k\rho k} R_{0,\lambda p\lambda q,\rho r\rho s} + \sum_{\nu} \Gamma_{\nu,\lambda i\rho k,\rho k\lambda i} R_{\nu,\lambda p\rho r,\rho s\lambda q} \Big) c_{\rho kr} c_{\rho ks} \quad (71)$$

where we have used the definition for the closed shell potential

$$G_{\lambda pq} = H_{\lambda pq} + \sum_{\rho rs} D_{\mathscr{C},\rho rs} P_{\lambda pq,\rho rs} \quad (66a)$$

as well as the definition of the core density matrix

$$D_{\mathscr{C},\lambda pq} = \sum_{i \subset \mathscr{C}} d_\lambda c_{\lambda pi} c_{\lambda qi} \quad (72)$$

and the total density matrix

$$D_{\mathscr{T},\lambda pq} = D_{\mathscr{C},\lambda pq} + \sum_{i \subset \mathcal{O}_\lambda} \gamma_{\lambda ii} c_{\lambda pi} c_{\lambda qi} \quad (73)$$

Using these definitions we can also write down the expectation value for the total energy in the convenient form

$$E = E_{\mathscr{C}} + \sum_{\lambda i \subset \mathcal{O}} \sum_{p,q} \tfrac{1}{2} (F_{\mathcal{O},\lambda ipq} + G_{\lambda pq}) c_{\lambda pi} c_{\lambda qi} \quad (65a)$$

with the closed shell energy

$$E_{\mathscr{C}} = \tfrac{1}{2} \sum_{\lambda pq} D_{\mathscr{C},\lambda pq} (G_{\lambda pq} + H_{\lambda pq}) \quad (67a)$$

Equations (64) through (73) are the detailed and complete set of working equations for the general open shell HF calculation using symmetry adapted orbitals. In the case of closed shell systems (64), (66), (67), (70), and (72) are sufficient with the Fock equation, (68), having no off-diagonal Lagrangian multipliers on the right-hand side.

E. Solving the SCF Equations

It may be appropriate to add some comments on the mathematical procedures of solving the Fock equations, in particular for cases where (*a*) the overlap matrix S is not the unit matrix and (*b*) there are off-diagonal Lagrangian multipliers. To do this we drop the symmetry indices again, in order to obtain a more compact notation. Extension to the general case with specific inclusion of symmetry should present no difficulty, since it merely means solving the same problem of smaller dimension several times, once for each symmetry.

1. *The Eigenvalue Problem with an Overlap Matrix*

Solution of the CI problem

$$\mathbf{HC} = \mathbf{CE} \tag{74}$$

or the HF equation for a closed shell case with the unit matrix as overlap matrix

$$\mathbf{Fc} = \mathbf{c}\epsilon \tag{75}$$

with given **H** or **F** matrix presents little difficulty, since these are standard eigenvalue problems, since **H** and **F** are symmetric. The unitary (orthogonal) matrices **C** and **c**, which contain the expansion coefficients as column vectors, are obtained by standard diagonalization techniques of the **H** of **F** matrix. If the dimensions of the matrices are less than 30 the Jacobi diagonalization method is efficient and adequate. For matrices of larger dimensions it is advisable to use more efficient techniques, for example, Householder tridiagonalization followed by the QL algorithm or bisection if only part of the eigenvalues and eigenvectors are required. Such routines are standard in most computation centers. In CI calculations, however, where it is common that **H** is too large to fit into the fast memory of a computer it is necessary to resort to iterative methods, which are efficient only for the evaluation of a few extreme, highest or lowest, eigenvalues. Fortunately it is generally adequate in these cases to compute only a few of the lowest roots.

The solution to the Fock equation

$$\mathbf{Fc} = \mathbf{Sc}\epsilon \tag{76}$$

where **S** is not the unit matrix presents only little extra difficulty. All one needs to do is obtain a vector set $\mathbf{c}_0$ forming a unitary matrix in the space with **S** as a metric, that is, such that

$$\mathbf{c}_0^{+}\mathbf{S}\mathbf{c}_0 = \mathbf{1} \tag{77}$$

is satisfied. In general such a vector set is already available since it had to be used in the construction of the Fock matrix; alternatively it can be obtained by an orthonormalization procedure, for example, Schmidt orthonormalization. One then forms

$$\mathbf{F}' = \mathbf{c}_0^{+}\mathbf{F}\mathbf{c}_0 \tag{78}$$

and solves

$$\mathbf{F}'\mathbf{c}_1 = \mathbf{c}_1\epsilon \tag{79}$$

using conventional matrix diagonalization techniques. We obtain

$$\mathbf{c} = \mathbf{c}_0\mathbf{c}_1 \tag{80}$$

as the desired eigenvector matrix, which clearly satisfies (76). The eigenvalues of (79) are identical to those of (76).

2. *The Fock Equations with Off-Diagonal Lagrangian Multipliers*

The Fock equation to be solved in the general open shell case is

$$\mathbf{F}_i\mathbf{c}_i = \mathbf{S}\sum_j \mathbf{c}_j\epsilon_{ji} \tag{81}$$

from which the off-diagonal Lagrangian multipliers may be obtained as

$$\varepsilon_{ji} = \mathbf{c}_j{}^+\mathbf{F}_i\mathbf{c}_i \tag{82}$$

Here $\mathbf{c}_i$ is the column vector of the expansion coefficient of orbital i.

Using (82) for the off-diagonal multipliers and subtracting them on the left of (81) we obtain

$$\left(\mathbf{F}_i - \sum_j{}' \mathbf{S}\mathbf{c}_j\mathbf{c}_j{}^+\mathbf{F}_i\right)\mathbf{c}_i = \mathbf{S}\mathbf{c}_i\epsilon_{ii} \tag{83}$$

where the total matrix in parentheses is nonsymmetric. It can be made symmetric, however, by adding the terms $\mathbf{F}_i\mathbf{c}_j\mathbf{c}_j{}^+\mathbf{S}$ giving

$$\left[\mathbf{F}_i - \sum_j{}' (\mathbf{S}\mathbf{c}_j\mathbf{c}_j{}^+\mathbf{F}_i + \mathbf{F}_i\mathbf{c}_j\mathbf{c}_j{}^+\mathbf{S})\right]\mathbf{c}_i = \mathbf{S}\mathbf{c}_i\epsilon_{ii} \tag{84}$$

The added terms are obviously zero for the solution vectors. One could now define a new Fock matrix equal to the square bracket of (84) and treat this as conventional eigenvalue problem. This is not appropriate, however, since we have not ensured that the solutions yield

$$\varepsilon_{ji} = \varepsilon_{ij} \tag{85}$$

as they should. This can be achieved by using instead of ε_{ji} of (82) an average value

$$\varepsilon_{ji} = \varepsilon_{ij} = \tfrac{1}{2}(\varepsilon_{ji} + \varepsilon_{ji}) = \tfrac{1}{2}\mathbf{c}_i{}^+(\mathbf{F}_i + \mathbf{F}_j)\mathbf{c}_j \tag{86}$$

in deriving (83) and (84); thus we obtain

$$\mathbf{F}_i{}'\mathbf{c}_i = \mathbf{S}\mathbf{c}_i\varepsilon_{ii} \tag{87}$$

with the modified Fock matrix as

$$\mathbf{F}_i{}' = \mathbf{F}_i - \tfrac{1}{2}\sum_j{}' [\mathbf{S}\mathbf{c}_j\mathbf{c}_j{}^+(\mathbf{F}_i + \mathbf{F}_j) + (\mathbf{F}_i + \mathbf{F}_j)\mathbf{c}_j\mathbf{c}_j{}^+\mathbf{S}] \tag{88}$$

Equation (87) may now be solved conventionally, that is, constructing the Fock matrices from initial orbital coefficient vectors, obtaining new improved orbital vectors as the solutions of the Fock equations. This process is repeated until the input orbitals agree within a given threshold with the resulting orbitals. Only for these final, converged orbitals will the

Fock equations be truly satisfied, and for them the transition from (81) to (87) and (88) is correct.

With the redefinition of the Fock operator (88), the Fock equations for the general open shell case, (87), appear to be just as simple as the normal closed shell HF equations; this is misleading, however. The iterative solution of the open shell equations will in general not converge to the final self-consistent answer as easily as it will for the closed shell equations. It should be pointed out that, although (87) and (88) are general enough to be applicable to a specific MCSCF form of the wave function, for this case SCF convergence will be obtained by solving (87) only in very special cases, that is, if no γ_{ii} is much smaller than 1. Otherwise more sophisticated iterative techniques are required. These are beyond the scope of this article.

F. Determination of the Initial Density Matrix

Before concluding this section, it seems worthwhile to point out a small "trick" that has been demonstrated to yield rapid SCF convergence in large molecule, closed shell calculations. The initial construction of the Fock matrix requires an initial guess of the orbital expansion coefficients or alternately of the first-order reduced density matrix, D_{pq}, in the space spanned by the basis functions. Clearly, the better this guess is, the fewer iterations will be required to achieve a self-consistent solution. It turns out that it is easy, in most cases, to obtain a "good" initial guess of the molecular density matrix, whereas it is not as easy to obtain a "good" initial guess of the orbital expansion coefficients for the canonical HF orbitals.

In most molecular calculations the basis functions chosen on the various atoms will correspond to atomic basis functions in terms of which the atomic wave functions are known. Thus it will be easy to construct the total atomic density matrices $D_{A,pq}$ for these basis functions. $D_{A,pq}$ will have nonzero elements only for those p and q that label atomic functions on atom A. A "good" approximation as an initial guess to the total molecular density matrix is then given as

$$D_{pq} \simeq \sum_{A}^{N} D_{A,pq} \tag{89}$$

An extension of this to include symmetry explicitly is straightforward. It is obvious that the atomic density matrices, $\mathbf{D}_A$, should be evaluated for configurations assumed by the atom in the molecule and not for the atomic ground configurations; for example, for carbon the configuration $1s^2 2s 2p^3$ should be used and not $1s^2 2s^2 2p^2$. Clearly (89) provides only a rough approximation to the molecular density matrix, giving zero elements

for all interatomic parts of D_{pq}; that is, $D_{pq} = 0$ if χ_p and χ_q are on different atoms. Such matrix elements, however, may be quite large. That (89) provides nevertheless a "good" starting approximation of the first-order reduced density matrix for the construction of the Fock matrix can be seen in the fact that the neglected interatomic terms of D enter the Fock equations multiplied by two-electron integrals containing an overlap-charge distribution, thus being small.

It is possible to obtain a better initial approximation to the molecular density matrix, including the dominant interatomic parts. Here we take advantage of the fact that it is generally quite easy to guess the orbital coefficients of localized, chemical type orbitals in terms of atomic orbitals or molecular fragment orbitals. Thus we construct a matrix of orbital expansion coefficient vectors c_i for the occupied molecular orbitals in the space of the basis functions by (*1*) choosing just the atomic expansion coefficients with zeroes everywhere else for orbitals that are closed shell in the atom and stay closed shell in the molecule, or in general are not involved strongly in the bonding; and (*2*) forming the appropriate linear combination of any two sets of atomic orbital expansion coefficients for all those atomic orbitals linked by a bond (note that it may be appropriate here and under *1* to use atomic hybrid type orbitals). The rectangular matrix $\mathbf{c}$ with orbital vectors $\mathbf{c}_i$ thus obtained will describe localized molecular orbitals which are either singly or doubly occupied; however, these orbitals will not be orthonormal, but nearly so. The overlap matrix between them is given by

$$\mathbf{S}' = \mathbf{c}^{+}\mathbf{S}\mathbf{c} \tag{90}$$

where $\mathbf{S}$ is the normal overlap matrix of the basis functions.

A good approximation to the total molecular density matrix is now obtained as

$$\mathbf{D} \cong \mathbf{c}\mathbf{S}'^{-1/2}\mathbf{D}'\mathbf{S}'^{-1/2}\mathbf{c}^{+} \tag{91}$$

where $\mathbf{D}'$ is the density matrix in the space spanned by the localized MO's, that is, a diagonal matrix with 2 or 1 on the diagonal for the doubly and singly occupied molecular orbitals, respectively. The transformation with $\mathbf{S}'^{-1/2}$ is introduced for proper symmetric orthonormalization of the initial vectors describing the localized MO's.

The success of using the simpler approximation, (89), as an initial guess of the first-order reduced density matrix in the construction of the Fock matrix, and the fact that $\mathbf{D}$ of (91) will differ from $\mathbf{D}$ of (89) only in those parts directly connected with bond formation, points to an important characteristic; the first-order reduced density matrix, or parts thereof, in the space spanned by the atomic basis functions appears to be nearly

transferable from molecule to molecule. We make use of this transferability in the next section.

G. Bibliography to the CI and SCF Methods

The entire framework developed here for calculation of the electronic wave functions was based on the Born-Oppenheimer approximation[1] and the variational principle.[2] The proof that the Kth root of a secular equation, (12), is an upper bound to the true energy of the Kth state is given by Hylleraas and Undheim[3] as well as by MacDonald.[4] The theory leading to variational equations for the determination of the orbitals goes back to Hartree,[5] Fock,[6] and Slater.[7] This was cast into the basis function expansion form by Roothaan[8] and by Hall[9] for the closed shell case. The derivation of the analogous SCF equations for open shell systems has been presented by Lefebvre,[10] Roothaan,[11] Huzinaga,[12] and Birss and Fraga.[13] General computational procedures for solving the open shell HF problem are discussed by Roothaan and Bagus,[14] McWeeney,[15] and Hunt.[16] The shortcomings of the HF method, that is, the electron correlation problem, have been reviewed by Lowdin.[17] Computational methods mostly based on the CI formalism and the use of approximate natural orbitals are discussed in detail by Schaefer[18] in a survey of rigorous quantum mechanical results of electronic structure calculations. Natural orbitals introduced by Lowdin[19] have been reviewed recently by Davidson.[20] The use of reduced density matrices in electronic structure calculations is presented in detail by Lowdin[19] and in a review by McWeeney.[21] The MCSCF method, yielding SCF equations for the orbitals of a many configuration wave function goes back to Frenkel,[22] and has been developed extensively by Das and Wahl[23] and Hinze;[24] for a detailed bibliography see Gilbert.[25] The use of STF's as basis functions for the expansion of orbitals was first suggested by Slater.[26] Procedures to evaluate the required molecular integrals over STF's are described by Wahl[27] and Ruedenberg[28] and others; however, to date there are still no satisfactory (accurate and efficient) procedures available for the evaluation of three- and four-center integrals for nonlinear general geometry molecules. It is for this reason that GTF's and GLF's have been introduced by Boys[29] and Preuss,[30] respectively. With the advent of Gaussian basis functions, permitting the rapid calculation of large numbers of molecular integrals, the time-consuming step in such a calculation has become the SCF or CI calculation. This is also true if semiempirical integral approximations are used. This makes it necessary to use sophisticated and efficient algorithms for the determination of eigenvalues and eigenvectors. Useful in the SCF calculations is the Householder method followed by the QL algorithm, described

by Wilkinson[31] and Dubrelle;[32] in a large CI calculation the algorithms developed by Nesbet[33] and Shavitt[34] are highly efficient.

III. PSEUDOPOTENTIALS; FROZEN CORE AND FROZEN GROUP APPROXIMATION, MODEL POTENTIALS

The idea of a pseudopotential[35] approach is the fact that in the ab initio as in the semiempirical computation of the electronic structure of large molecules it is generally desirable and feasible to focus in the calculation on that part of the electronic structure of the molecule that is changing in bond formation or in excitation. Such a procedure is generally followed, for example, core shells are described with minimal basis sets only, or in the calculation of correlated wave functions only the valence shell is correlated, while the core is only SCF adjusted. We can go one step further, keeping the core or other parts of the wave function fixed, that is, keeping some of the orbitals unchanged in the SCF process, thus reducing the amount of necessary computational effort. It is clear that such a "frozen core" or "frozen group"[36] approximation will be satisfactory only if there is good reason to believe that the set of orbitals that is held constant would not change much if they were allowed to adjust freely in the process studied. What is actually important is that the density matrix part due to these orbitals would not change much. The observation presented at the end of the preceding section gives us confidence that this will hold in many cases to a quite good approximation. We now treat such a "frozen core" approximation, developing the corresponding "pseudo" potential exactly for the motion of the valence electrons, which are to be treated explicitly. As we will see, there is no difficulty due to the nonlocal exchange potential or due to the necessary orthogonality constraints, as long as we are working within the framework of a basis function expansion of the orbitals. For convenience of notation we do not use symmetry in the following derivation. The detailed inclusion of symmetry does not alter the argument presented.

Let us assume that we have a set of t basis functions $\{\chi_1, \chi_2, \ldots, \chi_t\}$ for the fixed and valence orbitals and the corresponding integrals

$$\begin{aligned} S_{pq} &= \langle \chi_p \mid \chi_q \rangle \\ H_{pq} &= \langle \chi_p \mid h \mid \chi_q \rangle \\ R_{pq,rs} &= \langle \chi_p \langle \chi_r \mid g \mid \chi_s \rangle \chi_q \rangle \end{aligned} \tag{92}$$

The basis set is now transformed to a new basis by a linear transformation

$$\{\kappa_1 \kappa_2 \cdots \kappa_k \kappa_{k+1} \cdots \kappa_t\} = \{\chi_1 \chi_2 \cdots \chi_t\}\mathbf{C} \tag{93}$$

where the first k functions $\{\kappa_1 \cdots \kappa_k\}$ correspond to the orbitals to be held fixed. It is necessary to require that these first k functions be orthonormal and that the functions $\{\kappa_{k+1} \cdots \kappa_t\}$ be orthogonal to them, that is,

$$\langle \kappa_i \mid \kappa_j \rangle = \delta_{ij} \qquad \text{for } i \text{ and } j \leqslant k$$

and

$$\langle \kappa_i \mid \kappa_j \rangle = 0 \qquad \text{for } i = 1, \ldots, k \quad \text{and} \quad j = k+1, \ldots, t \tag{94}$$

Thus the $t \times t$ quadratic matrix $\mathbf{C}$ can be thought of as being composed of a matrix $\mathbf{A}$ with t rows and k columns describing the fixed orbitals and a matrix $\mathbf{B}$ with t rows and $u = t - k$ columns describing the remainder of the basis function space such that

$$\mathbf{A}^+\mathbf{S}\mathbf{A} = \mathbf{1}(k \times k)$$

and

$$\mathbf{A}^+\mathbf{S}\mathbf{B} = \mathbf{0}(k \times u). \tag{95}$$

The column vectors of $\mathbf{A}$, $\mathbf{a}_i$, describing the orbital κ_i to be fixed should be obtainable directly from previous atomic or molecular group SCF calculations, for example, the vector corresponding to the K shell of an atom with zeroes at the positions of all other basis functions. The requirement that the $\mathbf{a}_i$ be orthonormal, to be achieved by a symmetric orthonormalization, should not modify these vectors significantly; otherwise the approximation of holding them fixed will not be good. The fixed orbitals κ_i, described by $\mathbf{a}_i$, will not be canonical HF type orbitals of the total molecular system; this is irrelevant, however, as long as they belong to a closed shell part of the total molecular wave function desired and are closed shell in the fragment. The requirement that the fixed orbitals, κ_1 through κ_k, are closed shell orbitals cannot be lifted easily.

We now define an effective one-electron potential for the motion of the valence electrons as

$$G_{pq} = H_{pq} + \sum_{rs}^{t} D_{\mathscr{F},rs}(R_{pq,rs} - \tfrac{1}{2}R_{ps,rq}) \tag{96}$$

where we have used the first-order reduced density matrix

$$D_{\mathscr{F},pq} = \sum_{i=1}^{k} 2a_{pi}a_{qi}$$

due to the fixed orbitals in the space spanned by the original basis functions χ.

The original t-dimensional basis space can now be reduced to u dimensions by transforming the matrices of integrals, (92) and (96), with $\mathbf{B}$ giving

$$\begin{aligned} \mathbf{S}' &= \mathbf{B}^+\mathbf{S}\mathbf{B} \\ \mathbf{H}' &= \mathbf{B}^+\mathbf{G}\mathbf{B} \\ \mathbf{R}' &= \mathbf{B}^+ \times \mathbf{B}^+\mathbf{R}\mathbf{B} \times \mathbf{B} \end{aligned} \tag{97}$$

The transformed integrals of the matrices $\mathbf{S}'$, $\mathbf{H}'$, and $\mathbf{R}'$ in the u-dimensional space spanned by the basis functions $\kappa_{k+1}, \ldots, \kappa_t$ can now be used in any conventional calculation for the determination of the energy and the orbitals for the valence electrons. If the total electronic energy is required, one needs to add the energy

$$E_{\mathscr{F}} = \tfrac{1}{2} \sum_{pq}^{t} D_{\mathscr{F},pq}(H_{pq} + G_{pq}) \tag{98}$$

due to the electrons in fixed shells. The orbital expansion vectors of the valence shells in the space spanned by the original basis functions can be obtained easily from the expansion vectors in the reduced space from

$$\mathbf{B}' = \mathbf{Bc} \tag{99}$$

where $\mathbf{B}'$ is the matrix of valence orbital expansion vectors in the original basis χ and $\mathbf{c}$ is the matrix of eigenvectors describing the valence orbitals in the reduced space $\{\kappa_{k+1}, \ldots, \kappa_t\}$.

The derivation above shows that if the wave function calculation is done in a basis function expansion form, and if one is willing to compute the integrals, (92), over all the basis functions, then the "frozen core" or "frozen group" approximation, that is, freezing some closed shell orbitals, presents little difficulty. There are no problems with nonlocal potentials due to exchange or the orthogonality requirement. The theoretical formulation is on solid ground, and significant computational effort can be saved in the determination of the valence orbitals and energy.

Furthermore, the derivation presented permits us to gauge the approximations necessary to treat the frozen part as a model potential,[37] if one desires to compute only those integrals over the basis functions necessary for the description of the valence orbitals. Let us now consider under what conditions we expect such a model potential approximation to be acceptable, and what the form of such a model potential should be. To do this we examine the matrix $\mathbf{C}$ of (93)

$$\mathbf{C} = \left(\begin{array}{c|c} \mathbf{A}_1 & \mathbf{B}_2 \\ \hline \mathbf{A}_2 & \mathbf{B}_1 \end{array}\right) \tag{100}$$

for which we have arranged the original basis $\{\chi\}$ such that the first k functions describe the orbitals to be frozen as well as possible. If $\mathbf{A}_2$ and $\mathbf{B}_2$ are zero or nearly so, then $\mathbf{S}'$ and $\mathbf{R}'$ will be equal to or nearly equal to the integrals calculated from the reduced basis functions $\{\chi_{k+1}, \ldots, \chi_t\}$ provided $\mathbf{B}_1 \cong \mathbf{1}$, which can always be achieved if $\mathbf{A}_2$ and $\mathbf{B}_2$ are nearly zero. $\mathbf{A}_2$ and $\mathbf{B}_2$ will be zero, or nearly so, if there is little or no overlap between the basis functions describing the frozen and valence part of the wave function, either because they are in different regions of space or

because they are of different symmetry, for example, σ and π shells in organic molecules.

Since a significant computational saving of a model potential approximation over the better founded frozen core approximation is obtained only when the two-electron integrals R can be computed directly only over the reduced basis $\{\chi_{k+1}, \ldots, \chi_t\}$, we conclude that the model potential approximation is good and useful only if $\mathbf{A}_2$ and $\mathbf{B}_2$ are small. Under these conditions it is also permissible to treat the potential due to the fixed (disregarded) electrons, defined explicitly in the second part of (96), as an empirical or semiempirical local potential. The exchange integrals in (96) represent the interaction of two overlap charge distributions between little overlapping "frozen" and "valence" type basis functions; thus they are small.

From this discussion it can be seen that the separate treatment of π electrons in large organic systems, where the σ skeleton is included via an empirical model potential, is justifiable, and so is the neglect of K shells in the various zero differential overlap methods. Furthermore, we would expect model potential approximations to hold when highly excited Rydberg types states are being calculated.

IV. BASIC SETS

In Sections II and III we have outlined the formalisms for the calculation of the electronic wave functions of molecules—wave functions which may or may not include electron correlation and which may represent ground as well as excited states. The practical implementation of these procedures is generally based on the expansion of the molecular orbitals, used in the construction of the wave function, in terms of basis functions. Thus the basis functions represent the fundamental building blocks of an electronic wave function. The accuracy and reliability of the results—energy or properties—attainable are ultimately determined by the adequacy of the basis set selected for such a calculation. If the basis set is inappropriate or inadequate for the faithful representation of the one-electron charge density, it will be impossible to obtain a good description of the molecular wave function, energy, and properties. Such a basis set inadequacy cannot be compensated for by any devices, such as the inclusion of electron correlation, and it may lead to erroneous results and predictions. The problem of a proper basis set selection cannot be avoided by simply using a very large, extensive basis set, which is overly adequate, trusting that the variational principle will select in the course of the calculation that part of the one-particle space necessary for the proper description of the charge density. Although the variational principle will do this, there

are two difficulties: (*1*) the basis may become nearly redundant, leading to numerical difficulties, and (*2*) more important, the computational effort necessary is roughly proportional to t^4, where t is the number of basis functions. Clearly this requires a compromise, in particular for large molecules, to arrive at an adequate basis set which is as small as possible, and for which the required computations can be carried out rapidly. To this end, we need to understand the qualitative principles which should determine such a compromise basis set selection.

A. Type of Function

We have to choose between the commonly used functions, that is, Slater-type functions, STF's of (18) and (19), Gaussian-type functions, GTF's of (20), and Gaussian-lobe functions, GLF's of (21). For reasons of accuracy the STF's centered on the various atoms of the molecule would be the most satisfactory. They resemble most closely the atomic functions; that is, it is possible to describe an atomic orbital well with only few STF's, having the proper cusp behavior at the nuclei,[38] and they can describe correctly the long-range exponential decay[39] of the one-particle functions. Being able to describe the atomic functions correctly in a molecular calculation is important, since the atoms in a molecule by and large retain their characteristic identity. Thus using STF's as the molecular basis functions would allow the use of the fewest such functions for a given desired accuracy. Unfortunately STF's have an important drawback; the evaluation of the required electron-electron repulsion integrals[27,28]

$$R_{pq,rs} = \langle \chi_p \langle \chi_r | \frac{1}{r_{12}} | \chi_s \rangle \chi_q \rangle \tag{101}$$

is quite cumbersome and time consuming for general geometry polyatomic molecules in particular, where the basis functions in (101) may be on three or four different centers. For linear molecules efficient algorithms and computer codes for the evaluation of these integrals have been developed,[40] and here STF's are the best basis functions to be used. However, for general geometry molecules the evaluation of these integrals is still too slow and cumbersome ($t^4/8$ such integrals must be evaluated in a calculation) and presents a severe bottleneck toward the use of STF's for such calculations. Recently significant progress has been made and some computer codes for the calculation of such integrals are available.[41,42] It appears, however, that for these cases it is still more efficient and effective to use GTF's or GLF's.

With the use of Gaussian-type basis functions, GTF's or GLF's, the evaluation of these integrals, one- as well as two-electron integrals, presents

little problem.[29,43] Efficient algorithms as well as computer codes exist,[44] algorithms that are most simple for the GLF's. However, it is not easily possible to describe with Gaussian-type functions the cusps of the wave function at the nuclei and the proper long-range tail behavior. Furthermore, it is generally necessary to use many more Gaussian functions than STF's in order to achieve comparable accuracy. Nevertheless it is still more efficient in general geometry molecules to use Gaussian basis functions rather than STF's. With this one sacrifices some accuracy, in particular for properties that depend critically on the wave function near the nuclei (Fermi contact term or field gradients at the nuclei, etc.), or on the long-range behavior of the wave function. In general, energies, and in particular energy differences, and molecular geometry can be obtained well using a limited number of Gaussian functions; also the electronic charge distribution can be represented well. This implies that molecular dipole and quadrupole moments can be obtained accurately.

Several (four to six) GTF's are required to represent the radial part of an atomic orbital well; even more GLF's are required, since it takes two GLF's to simulate the angular part of a *p*-type function and four to simulate that of a *d*-type function. In order to avoid using that many *primitive* basis functions in the wave function calculation, it has been found more efficient, and it has become customary, to form *contracted* basis functions[45]

$$\chi_p = \sum_v \kappa_v b_{vp} \tag{102}$$

where the κ_v's are the primitive GTF's or GLF's, while the χ_p's are the contracted basis functions in terms of which the actual energy and wave function calculation is carried out. The expansion coefficient b_{vp} are predetermined and fixed once and for all by either fitting the contracted functions to atomic STF's[46,47] or by optimizing them in detailed atomic or molecular fragment calculations.[48]

Using such contracted functions presents a significant advantage in the energy calculation, although some flexibility is lost. It is necessary only to compute all the one- and two-electron integrals over the primitive basis functions. These integral lists are then significantly reduced by transforming them to integrals over the contracted functions. Only these reduced integral lists need be used in the energy and wave function calculation, which leads to a significant saving in time, since all basis function integrals must be read for each SCF iteration. An additional computer time saving may be realized by the use of *adjoint* basis sets,[49] that is, single GTF's which roughly represent the contracted functions χ_p of (102). These adjoint basis functions can be used to estimate rapidly the magnitude of the integrals over the contracted functions. If such an

estimate results in a magnitude for an integral of order 10^{-8} or less, say, such an integral need not be computed at all, or if the estimate yields a magnitude of 10^{-4} or less, say, it is sufficient to compute this integral using the rough, adjoint basis functions only, to obtain directly the integral over the contracted basis functions. Such a process can yield quite dramatic time savings, in particular in the case of large molecules, with little or no loss of accuracy.

B. Basic Function Characteristics

In addition to the type of basis functions selected (STF, GTF, or GLF), it is important to choose the appropriate basis function characteristics, that is, whether they should be atomic *s*-, *p*-, or *d*-like, what the exponential factors ζ should be, and in the case of GLF's, where they should be centered. Also important here is how many basis functions will be necessary for truly quantitative results, and what is sacrificed if one chooses fewer.

We should realize first that an atom, becoming part of a molecule, by and large retains its identity, though in general its outermost orbitals, that is, valence orbitals, get polarized and contracted.[50] Second, the inner atomic shells, which are not involved in the changes due to bond formation, change little. They need not be described accurately if one does not desire properties near the nuclei but only differences in energy and charge distribution due to bond formation or valence electron excitation. With this in mind we can draw on the experience gained in accurate atomic calculations, where it is computationally feasible to variationally optimize all the basis function characteristics. Thus a good guide for the selection of atomic centered STF's are the nominal atomic basis sets by Bagus et al.[51] or the double zeta sets of Clementi.[52] In the case of atomic centered GTF's the contracted sets of Huzinaga[46] or Dunning[53] are about equivalent to Clementi's double zeta sets, as are the contracted GLF sets of Witten.[45] Frequently it will be possible to economize by representing the *K* or lower *L* shells by a more compact single zeta (single basis function) representation, while still maintaining double zeta flexibility in the atomic valence shells. For quantitative calculations it is also important to augment these atomic basis sets by polarization functions, that is, atomic functions with higher *l* value with roughly the same radial dependence as the atomic valence shell functions.[54] Atoms within molecules are clearly in strong electric fields and get polarized or, what is the same, they hybridize. Obviously, when going to larger molecules it will be necessary to use more and more restricted, almost minimal, basis sets. However, one should then be cautious, for some of the results may be erroneous and lead to wrong conclusions. For example, the barrier of inversion of NH_3 calculated without *d*-polarization functions on nitrogen is quite wrong.[55] Similarly,

dissociation energies calculated with minimal or almost minimal basis sets may be significantly too large, that is, even He_2 can become bound,[56] since atom one uses the basis functions on atom two to make up for some of its own basis set deficiency and vice versa, leading to a "*basis function bond.*"

Another caution is required when calculating excited states or negative ions. Here it will be necessary to augment the atomic basis sets by low exponent functions in order to allow properly for the expanded valence charge distribution. A guide for the specific low exponents to be chosen is the orbital energy of the highest occupied orbital, in that the smallest exponents describing the tail behavior should be

$$\zeta \cong \sqrt{-2\eta} \tag{103}$$

for STF's and the equivalent for GTF's.[57] Here η is the orbital energy of the highest occupied molecular orbital. This requirement for the tail behavior of the wave function indicates clearly that wave functions obtained for negative ions, where orbitals with positive orbital energy have to be occupied, should be taken with a grain of salt.

In the case of GLF's it has been found advantageous and effective in some calculations to use in addition to the functions describing the atomic orbitals some GLF's centered in the bonding region.[58] Here it appears quite difficult to predict where such functions should be centered, and what their exponential coefficients should be. These parameters can be found only by optimization in calculations of molecular fragments—a time-consuming and computationally cumbersome nonlinear optimization.[59] Once the parameters are fixed for particular fragments they can be used in large molecule calculations.[60]

V. DISCUSSION OF APPROXIMATE MOLECULAR WAVE FUNCTIONS

It is well known that the true electronic molecular wave function can be approximated arbitrarily close by the methods described above. However, this requires the use of quite extensive basis sets and the introduction of correlation corrections using CI[18] or MCSCF[23,24] methods or perturbation methods[61] not described here. Clearly such computations are quite extensive and at present suitable only for small molecules. In the calculation of medium or large molecules one frequently has to be satisfied with less extensive basis sets and uncorrelated HF type wave functions. In particular, in medium molecules, it is possible to introduce some of the important correlation corrections using explicitly a limited MCSCF[23] or CI[18] approach or a semiempirical correlation correction.[62] In fact, it can

be expected that these important corrections will also become feasible in large molecule calculations. To data most ab initio calculations on large molecules are restricted to the HF approximation, and we should discuss what such approximate wave functions can yield and what their shortcomings are.

It is well known that closed shell HF wave functions satisfy Brillouin's theorem,[63] that is, energy matrix elements between the HF single determinantal function Φ_{HF} and determinants obtained by replacing a spin orbital in Φ_{HF} by another spin orbital are zero. This single replacement function is most succinctly expressed with the use of creation and annihilation operators a^+ and a.[64] Thus Brillouin's theorem may be written as

$$\langle a_k^+ a_i \Phi_{HF} | \mathscr{H} | \Phi_{HF} \rangle = 0 \tag{104}$$

where $a_k^+ a_i \Phi_{HF}$ represents the determinant obtained by replacing spin orbital i in Φ_{HF} by spin orbital k. For open shell HF wave functions or for MCSCF functions, Ψ, a similar extended theorem holds;[65] that is,

$$\langle (a_k^+ a_i - a_i^+ a_k)\Psi | \mathscr{H} | \Psi \rangle = 0 \tag{105}$$

A consequence of Brillouin's theorem is that all properties described by one-electron operators will be obtained from HF functions good through second order in perturbation theory.[66] Thus the total electronic charge distribution, electric moments, and in particular, dipole moments, etc., will be represented well by HF type functions. An exception to be mentioned here occurs when there is a nearly degenerate double excitation which would mix significantly with the HF function.[67] Then a few-configuration MCSCF function would be required to obtain good one-electron properties.

The failure of HF functions to describe bond dissociation adequately is also well known. In general, within the restricted HF molecular orbital model, bonds dissociate not to neutral atoms or molecular fragments but to a linear combination of neutral and ionic fragments. In general it would take one extra configuration per bond in an MCSCF type function to yield correct bond dissociation, but even then the bond dissociation energy would not be reproduced well. The quantitative description of bond dissociation energies in ab initio calculation requires at least the incorporation of the extra electron correlation energy due to the extra electron pair formation in the bond.[68] This is in general still out of the reach of present-day computational methods for large molecules. Only in the case of small and intermediate type molecules can such quantitative results be obtained to date. However, it is possible to obtain within the HF model heats of reactions for some special rearrangement reactions, as long as the number of closed shells in reactants and products is the same.

For example, for the reaction of the type

$$H_2C{=}O + 2H_2 \rightarrow CH_4 + H_2O$$

or

$$H_2C{=}C{=}O + CH_4 \rightarrow H_2C{=}CH_2 + H_2C{=}O$$

ΔH is adequately reproduced within the HF model;[69] this is due to the fact that the correlation errors in products and reactants are about equal, thus cancelling in the determination of the heat of reaction. Many reactions of interest satisfy the above criteria.

Ionization potentials of the valence shells can be obtained quite reliably and easily from HF results using Koopmans' theorem;[70] that is,

$$I_i = -\eta_i \tag{106}$$

Where I_i is the ionization energy necessary to remove an electron from the canonical HF orbital φ_i, and η_i is the corresponding orbital energy in our normalization of the Fock equations, (37) and (38), say,

$$\eta_i = \frac{\epsilon_{ii}}{\gamma_{ii}} = \tfrac{1}{2}\epsilon_{ii} \tag{107}$$

where γ_{ii} is the orbital occupation number. Alternately one may compute the HF wave functions and energies for the corresponding neutral and ionic species and obtain the ionization energy as the difference of the two energies for neutral and ion obtained. However, ionization energies obtained from such an energy difference, in addition to requiring more computational work, generally are in poorer agreement with experimental ionization energies than those obtained using Koopmans' theorem. For closed shell neutral systems they are generally too low by 1 or 2 eV, while Koopman's ionization energies will agree with experiment with an error of about 0.5 eV. The reason for this is quite well understood, being a cancellation of two errors of about equal magnitude but opposite sign. The orbital energy of a closed shell HF equation is exactly equal to the difference of the HF energy of the neutral system minus the energy calculated for the positive ion, with the HF orbital functions of the neutral system. Since the positive ion energy is calculated using the neutral system functions, this energy will be larger than the HF energy for the positive ion. This error is compensated by the fact that in general the correlation error of the neutral system, having one more electron, is greater than that in the positive ion. This cancellation of errors is nearly perfect for valence electron ionization energies. More recently there has been a significant interest in lower K and L shell ionization energies, due to the availability of X-ray photoelectron (ESCA) spectroscopic data. Here the HF error frequently is much larger, since the change in the field due to a missing

core electron is quite sizable. Thus the HF error in using Koopmans' theorem is no more cancelled by the difference in the correlation error, and one obtains ionization energies that may be several volts too large if one uses (106).[71] Here the energy difference method is much more reliable. However, in either method ESCA chemical shifts should be represented well, although the absolute ionization energies are not quantitative.

Using the simple HF model to predict electron excitation energies is in general not reliable. This is particularly true if a *virtual* orbital is used to represent the extra excited state orbital in the construction of the excited state CSF with which the excited state energy is computed. To explain this let us refer back to the simple closed shell HF equations, (37) and (38), which determine the lowest $n/2$ orbitals of an n (even) electron system or, what is the same, the orbital expansion coefficient vectors $\mathbf{c}_i$ for $i = 1, \ldots, n/2$ in (37a). When diagonalizing the Fock matrix of (37a) and (38a) one obtains in addition $t - n/2$ orbital expansion vectors, which define the virtual or unoccupied orbitals. Here t is the number of basis functions used, that is, the dimensionality of the Fock matrix. Little or no physical meaning can be attributed to these virtual orbitals or to the corresponding diagonal Lagrangian multipliers. This is so for two reasons: (*1*) the virtual orbitals describe roughly an extra electron moving in the field of the n closed shell electrons of the neutral ground state and should generally be unbound (positive ε_{ii}'s). This is definitely not what one wants for an excited state orbital, which should describe an electron moving in the field of $n - 1$ electrons. It is also not adequate for the representation of the negative ion, since here a significant rearrangement of the underlying closed shell would be required, as well as the inclusion of correlation effects, in order to get the extra electron bound. (*2*) We can define a projection operator

$$P = \sum_{i=n/2+1} |\varphi_i\rangle\langle\varphi_i| \tag{108}$$

in the space spanned by the virtual orbitals by summing only over the empty orbitals. Almost any operator, guarded left and right by such a projection operator, can now be added to the Fock operator of (37) without affecting the essential solutions of the Fock equation,[72] that is, the occupied orbitals and their energies. Specifically

$$\{F + PQP\}\varphi_i = \varepsilon_{ii}\varphi_i \tag{109}$$

has the same essential solutions as (37), although Q, which due to P acts on the virtual orbitals only, may be almost any operator. Thus using such an addition to the Fock operator the virtual orbitals may be modified at will as long as they stay orthogonal to the occupied orbitals. Clearly the

same argument can be used in the open shell HF case and in the case of the Fock equations in their basis function expansion form.

It is now possible to use this freedom discussed above to advantage by formulating the free operator Q such that $F + Q$ resembles closely the open shell Fock operator one would obtain were one to determine the excited state orbital directly from an open shell HF treatment for the excited state alone.[73,74] An excited state orbital obtained in this way together with all the ground state orbitals will have a physical meaning and the excited state CSF constructed from it will be a good representation of the excited state, provided the excited state can be described well by a single electron excitation. This process corresponds essentially to the determination of the extra excited state orbital in the frozen n-1 electron core described by the ground state orbitals, and one could use in such a calculation the integral or Fock matrix transformations described in Section III.[75] Since in such calculations the excited state wave function will not be of HF quality, while the ground state wave function is (provided the basis is adequate), and since in general the correlation error in the ground state will be larger than in the excited state, one would expect again a cancellation of errors as we had for the calculation of ionization potentials using Koopmans' theorem. Obviously, all this will be true only if the basis set used in the calculation is adequate, which may require particular care in the description of the excited states since here the basis functions of the composite atoms may not be a satisfactory guide. Also required is that the excited state can be represented well by a single electron excitation with little core rearrangement. How general this is for molecular excited states still remains an open question, deserving considerable more study and experience. This will require some extensive CI studies of such systems. Unfortunately, to date only a few ab initio CI calculations have been carried out for the description of excited states of larger molecules. This is in marked difference to the large number of excited state CI calculations which have been reported for conjugated π-electron systems using the semiempirical Pariser-Parr-Popel method, calculations which appear to agree with the available experimental information remarkably well.

In this context we should mention also the recent results obtained for excitation energies and transition probabilities of some large molecules[76] using the random phase approximation (RPA).[77] These ab initio results are in excellent agreement with the available experimental information, although in the case of ethylene, there appears to be a discrepancy between the RPA representation[76] of the V-N transition and that obtained by using a more conventional HF and extended CI approach[78] for the description of this transition. Since the RPA approximation should be related to a limited CI representation, though the detailed correspondence must be

demonstrated, one would suspect that the excellent agreement with experiment found for the RPA results is due to some fortuitous or systematic cancellation of errors. This question certainly appears worth investigation.

VI. LIMITATION OF AB INITIO METHODS

It is clear from all the preceding sections that so far we have biased this review toward ab initio calculations of molecular wave functions and molecular properties. This was done deliberately, because one can expect that in the future larger and larger molecules will be dealt with using ab initio methods, since only then will one obtain unambiguous and reliable molecular information, where the effects due to the approximations made (Born-Oppenheimer, nonrelativistic, HF approximation) are well understood and tractable. This is not to say that semiempirical methods are not of great practical value; however, it will always be necessary to calibrate the results obtained from semiempirical calculations against some experimental and ab initio results in order to assess the reliability of the predictions obtained. We should now assess up to what limits in molecular size ab initio calculations have been and can be carried out, and what procedures we expect to extend these limits in the future.

In an all-electron ab initio calculation using the methods outlined in Section II, together with the basis function expansion form of the orbitals, it will be necessary to use at least as many basis functions as there are electrons, that is, minimal basis set representation. One can choose even fewer, if one uses subminimal basis sets,[59] that is, one GLF centered in the bonding and lone pair regions to describe an electron pair. This idea of floating Gaussian subminimal basis sets has been introduced by Frost[59] and applied successfully by Christoffersen[79] to quite large molecules. Thus the smallest number of basis functions, t, which can be used in an all-electron ab initio calculation will be equal to n or $n/2$, where n is the number of electrons. The number of one-electron integrals that must be computed will be of order $t^2/2$ or proportional to n^2, while the number of two-electron integrals to be computed will be $t^4/8$ or proportional to n^4. In the case of a subminimal basis this becomes $n^4/128$. Using $n = 100$, well within our large molecule limit, we need to compute about 10^6 to 10^8 two-electron integrals, a computation that requires in the case of GTF's or GLF's on today's fastest computers (CDC 7600 or IBM 360/195) about 1 min to 1 hr; this is quite feasible. Much more limiting, in fact, is the peripheral storage space required to retain these integrals. Another saving feature is that in large molecules the basis functions centered far apart do not overlap significantly; thus all two-electron integrals involving

such an overlap-charge-distribution will be so small (less than 10^{-8} a.u.) that they need not be calculated or stored. Making use of this feature it is found that for molecules larger than benzene the number of two-electron integrals that need to be computed and stored is proportional to n^3 rather than n^4, resulting in a dramatic saving.[80] For example, a bench mark calculation of the cytosine-guanine complex ($C_9N_8O_2H_{10}$, i.e., $N = 29$, $n = 136$) using 105 contracted Gaussians, contracted from 334 primitive GTF's, required about 200 hr of IBM 360/195 time.[81] Here approximately 1.5×10^7 integrals over the contracted functions had to be constructed from approximately 1.5×10^9 integrals computed over the primitive functions. The same calculation using the adjoint basis, and not computing and storing the small integrals, required only 2 hr on the same computer,[80] a saving by a factor of 100.

From the above it can be seen that to date all-electron, ab initio calculations are feasible for systems up to about 200 electrons using minimal basis sets, and up to about 400 electrons if subminimal basis sets are used. This is true, however, only if all the above-mentioned savings are used and incorporated into the computer programs, which appears likely in the near future. That the number of integrals which need to be computed and stored in the case of large molecules is proportional to n^3 and not n^4 is important and should be kept in mind. It should be possible to treat even larger molecules using model potentials or the frozen core, frozen molecular fragment approximation discussed in Section III. Such approximations are being investigated and in part used successfully in ab initio type calculations.[82,83]

In order to show the extent to which ab initio methods can and have been applied to the electronic structure calculation of large molecules, we present in Table I an annotated list of some representative calculations. This list is by no means complete, for a complete tabulation would require several hundred entries; the examples included here are arbitrarily selected to illustrate what can be done to date. More comprehensive tabulations are available;[83,96] unfortunately they are somewhat out of date. It can be seen from Table I, and from the more comprehensive tables, that most large molecule calculations are restricted to a closed shell HF representation of the molecular ground states. It would certainly be worthwhile to explore more extensively the description of excited states and the feasibility of the inclusion of electron correlation effects, at least as a semiempirical correction.[97] However, it would be more satisfying if efficient ab initio methods could be found to introduce at least the most important correlation effects into such large molecule calculations without undue increase in the computational effort required. Methods based on the MCSCF[23–25] procedure, similar to the OVC[23] method, appear to be the most promising.

TABLE I

System	Ref.	N	n	Method; Type of Study
Trisubstituted methane CHXYZ; X, Y, Z = CH_3, NH_2, OH, F	84	5–14	34	HF, minimal STF's expressed in terms of contracted GTF's; ground state nuclear and electronic structure
Disubstituted ethane XCH_2-CH_2Y; X, Y = CH_3, NH_2, OH, F	85	8–14	34	HF, minimal STF's expressed in terms of contracted GTF's; ground state nuclear and electronic structure
$F^-\cdot 2H_2O$	86	7	30	HF, double zeta type contracted GTF's; ground state interaction energy surface
$NH_3\cdot H_2O$	87	7	20	HF, double zeta type contracted GTF's; ground state interaction energy surface
$Na^+\cdot (H_2O)_n$; $n = 1$ to 6	88	4–19	20–70	HF, double zeta type contracted GTF's; ground state interaction energy surface
$NH_2 + C_2H_4$	89	9	25	HF plus CI, extended contracted GLF's; reaction energy surface
Cytosine-guanine complex, $C_9N_8O_2H_{10}$	90	29	136	HF, minimal contracted GTF's; ground state energy, electronic structure, hydrogen bonds
Bicyclo[2.1.1]hexane, C_6H_{10}	91	16	46	HF, minimal contracted GTF's; ground state energy and electronic structure
Nickel tetracarbonyl, $Ni(CO)_4$ and $Ni(CN)_4^{2-}$	92	9	74	HF, minimal contracted GTF's; ground and excited states, ionization energies
Benzene and naphthalene isomers	93	12	42	HF, subminimal floating GLF's; molecular fragment; relative ground state stability
Acetylcholine, $C_7H_{16}NO_2$	79	26	81	HF, subminimal floating GLF's; molecular fragment; ground state electronic and geometric structure
Butadiene, C_4H_6	94	10	30	HF plus CI, extended, contracted GLF's; electronic and geometric structure of ground and excited states
HCN, FCN, HCF, HNO, FNO	95	3	14–24	HF plus CI, minimal STF's expressed in terms of contracted GTF's; excited states, geometric and electronic structure

VII. SEMIEMPIRICAL METHODS

As we have demonstrated above, ab initio methods are capable and efficient enough to permit the calculation of molecular wave functions of large molecules. However, computer programs for such calculations are quite complicated in order to do these calculations efficiently and manage the large amounts of data (integrals) that must be handled. In most of the calculations performed on large molecules to date semiempirical methods are used. Many of these semiempirical methods are based on the zero differential overlap approximation[98] in order to reduce the number of two-electron integrals. Notable examples are complete neglect of differential overlap (CNDO),[98] intermediate neglect of differential overlap (INDO),[99] modified INDO (MINDO),[100] and partial neglect of differential overlap (PNDO).[101] Although these methods are quite popular, permitting rather routine calculations, we do not discuss them in detail here, since they have been reviewed recently[102] and a detailed presentation of CNDO and INDO[103] as well as MINDO[104] is available in monograph form.

The zero differential overlap approximation

$$\chi_p(1)\chi_q(1)d(1) = 0 \tag{110}$$

if $p \neq q$ is used throughout in CNDO, whereas in INDO and MINDO the interactions of exchange charge distributions on the same center are retained, and only the two-center charge distributions are neglected as is the interaction of a one-center exchange charge distribution with the charge distribution on another center.

Using CNDO throughout, the integrals over the basis functions required for a molecular orbital calculation become

$$S_{pq} = \langle \chi_p \mid \chi_q \rangle \cong \delta_{pq} \tag{111}$$

$$\begin{aligned} H_{pq} = \langle \chi_p | \, h \, | \chi_q \rangle &= \frac{-1}{2} \langle \chi_p | \, \nabla^2 \, | \chi_q \rangle - \sum_A Z_A \langle \chi_p | \frac{1}{r_A} | \chi_q \rangle \\ &\cong -\tfrac{1}{2} \langle \chi_p | \, \nabla^2 \, | \chi_q \rangle - \delta_{pq} \sum_A Z_A \langle \chi_p | \frac{1}{r_A} | \chi_p \rangle \\ &= T_{pq} - \delta_{pq} \sum_A V_{A,p} \end{aligned} \tag{112}$$

and

$$R_{pq,rs} = \langle \chi_p \langle \chi_r | \frac{1}{r_{12}} | \chi_s \rangle \chi_q \rangle \cong \delta_{pq} \, \delta_{rs} \langle \chi_p \langle \chi_r | \frac{1}{r_{12}} | \chi_r \rangle \chi_p \rangle = \delta_{pq} \, \delta_{rs} \gamma_{pr} \tag{113}$$

Thus the number of integrals, which in an ab initio calculation is proportional to t^4 (where t is the number of basis functions), is now proportional to t^2, a drastic reduction. In addition only the valence electrons are

considered explicitly; the screening due to the neglected core electrons is incorporated simply by using an effective Z for Z_A in (112).

Using these approximations the closed shell Fock equations become

$$Fc = c\varepsilon \tag{114}$$

with the Fock matrix elements

$$F_{pq} = 2T_{pq} - 2\,\delta_{pq}\sum_{A} V_{A,p} + 2\,\delta_{pq}\sum_{r} D_{rr}\gamma_{pr} - D_{pq}\gamma_{pq} \tag{115}$$

Clearly the approximation used, (110), is quite drastic and it results in the requirement of an additional approximation; that is, integrals involving atomic p orbital charge distributions must be treated as if these charge distributions were spherical and the same for (p_xp_x), (p_yp_y), and (p_zp_z) and similarly for d orbitals.[98] This is necessary in order to make the results obtained independent of the choice of the coordinate system, *space invariance*. Another approximation frequently used of *hybridization invariance*,[98] that is, treating s and p charge distributions identically, is not essential. Using these necessary approximations, the problem is simplified even further since now many of the one- and two-electron integrals become identical, and if they are to be evaluated directly, they become simpler, since one deals with s-type charge distributions only.

In the light of the drastic approximations made, it is in general not necessary, and in fact not advisable, to compute the required integrals explicitly. The success of these semiempirical methods rests on the practice of approximating these integrals and treating some of them, in particular the T_{pq}'s, as empirical parameters determined either from atomic or molecular experimental data. With this one hopes to account for the neglect of the large number of integrals and even for the inclusion of some correlation correction, although the calculations are carried out within the HF approximation. Several parametrization schemes have been proposed[105–108] and are commonly used to obtain successfully molecular ground state information, such as ionization potentials, bond angles and distances, dipole moments, force constants, NMR and ESR data, and even heats of formation. A special parametrization of CNDO coupled with a limited CI expansion also yields information about excited states.[108]

The success of these methods cannot be denied, and large amounts of useful molecular information have been obtained with them. In fact CNDO type calculations of medium and large molecules have become rather routine; the required computer programs are simple and available,[109] and the computations are quite fast. However, the general belief should be tempered that for truly large molecules only CNDO type methods are applicable, whereas ab initio methods are not because they

would require more computational effort, orders of magnitude higher. It is true that in the CNDO methods the number of integrals required is only proportional to n^2 (where n is the number of valence electrons considered explicitly), whereas in ab initio calculations this number is proportional to n^3 (not n^4; see Section VI). Also CNDO type integrals are obtained much more easily, since they are frequently approximated via simple formulas. In CNDO, as in ab initio methods, it is required to diagonalize the Fock matrix, the dimension of which is proportional to n. The number of computer operations in such a diagonalization is proportional to n^3 even in the best algorithms. Thus in CNDO, as in ab initio calculations of large molecules, the computational effort required is inherently proportional to n^3. It is the proportionality factor which is about an order of magnitude smaller for CNDO type calculations, which makes them more attractive, in addition to the simplicity of the required computer programs relative to those of ab initio type programs, which need to be carefully optimized to yield the required efficiency outlined in Section VI.

Very similar to this comparison between ab initio and CNDO type methods is the comparison between CNDO and extended Hückel (EHT)[110] type methods. Here all electron interaction terms are neglected the the Hückel matrix (similar to the Fock matrix) is constructed using empirical values for all one-electron integrals.[111] How instructive and useful results can be obtained with such a crude approximation is amply demonstrated by the Woodward-Hoffman rules derived from EHT type calculations.[112] But again the computational effort required is proportional to n^3, since the Hückel matrix needs to be diagonalized. To be sure considerable savings could be realized were one to take advantage of the sparseness of the matrix; this is not easy however, and not being done in general. Again it is only the simplicity of the required computer programs and the proportionality factor of the n^3 term, which makes EHT calculations easier and faster than CNDO type calculations. The Hückel matrix needs to be diagonalized only once, while it is necessary to diagonalize the Fock matrix of a CNDO calculation several times, once per SCF iteration, or about 10 times in a calculation, unless one has a very good guess of the density matrix (see Section IV).

Two more approaches, which are halfway between ab initio and semiempirical, have been advocated and applied to the calculation of molecular wave functions of large molecules. One is based on the systematic approximation of all two-electron integrals using the Mulliken[113] or Ruedenberg[114] approximations (or a combination thereof),[115–117] or by using an approximate Gaussian representation of the basis functions to calculate these integrals rapidly. We need to say little here about this approach since it is quite similar to a direct ab initio calculation using contracted GTF's

together with the idea of an *adjoint* basis (with quite large thresholds) discussed in Section VI.

The other approach uses the self-consistent-field $X\alpha$ scattered wave method[118,119] (SCF-$X\alpha$-SW), which is an extension to molecules of the scattering wave method in solids.[120] The SCF-SW-$X\alpha$ method has little resemblance to the conventional MO approaches presented above, and we give only a sketchy outline of its salient feature here, since the method has been reviewed recently.[121] The space for the molecular wave function (charge distribution) is divided into three regions: (*1*) *atomic*—regions of nonoverlapping (touching) spheres centered at and surrounding each atom; (*2*) *interatomic*—a region spherically surrounding and enclosing all the atomic spheres; and (*3*) *extramolecular*, the region outside sphere II.

An approximation to the one-electron Schrödinger equation (Fock-like equation)

$$\{-\tfrac{1}{2}\nabla^2 + V_c + V_{X\alpha}\}\varphi_i = \varepsilon_i\varphi_i \tag{116}$$

is solved to obtain the orbitals φ_i and the orbital energies ε_i. Here V_c is the coulomb potential defined by the atomic nuclei and the total electronic charge distribution

$$\rho(1) = \sum_i \gamma_{ii}\varphi_i^*(1)\varphi_i(1) \tag{117}$$

The exchange potential $V_{X\alpha}$ is expressed using the Slater $X\alpha$ statistical exchange approximation[122] separately for α and β spin electrons

$$V_{X\alpha}(1, s) \cong -3\alpha\left[\frac{3}{4\pi}\rho(1, s)\right]^{1/3} \tag{118}$$

where s stands for α or β, the charge distribution $\rho(1, s)$ being calculated as in (117) but using only the α or β spin orbitals in the sum. In addition to the $X\alpha$ approximation the potentials V_c and $V_{X\alpha}$ in (116) are spherically averaged in regions 1 and 3 and assumed constant in region 2. With this (116) can be solved easily in the individual regions. In regions 1 and 3 the angular and radial part can be separated and the radial Schrödinger equation can be integrated numerically as in the case of free atoms. In region 2 the solutions are incoming and outgoing scattered waves, which are matched at the boundaries to the solutions in regions 1 and 3.

The ability to separate coordinates in regions 1 and 3 together with the analytical scattered wave type solutions in region 2, due to the averaging of V_c and $V_{X\alpha'}$, allows one to perform quite rapidly computations on large polyatomic molecules. No large lists of molecular integrals need be computed, nor is it necessary to diagonalize large matrices. This method has been applied to a number of intermediate[123] and large molecules,[124] and the results indicate that the orbital energies obtained can be matched

well with observed ESCA and photoelectron spectroscopy data. An advantage of this method is that molecules are described such that they dissociate correctly to neutral atoms, although a Hartree-Fock type of approximation is used. A disadvantage is that molecular geometries are predicted unreliably at best; for example, H_2O is obtained as being linear,[125] unless extra region 1 type spheres are introduced to describe the lone pairs. Also dissociation energies and force constants are obtained significantly too low, and for excitation energies one can obtain only configuration averages of several multiplets. It appears reasonable that the $X\alpha$ method should reproduce molecular electric moments correctly, since the total charge distribution should be adequately described. Unfortunately, to date no such moments, dipole or quadrupole, have been reported in the literature; such results would be quite interesting. Presently the method appears promising, since it seems to permit the rapid calculation of approximate molecular wave functions of large molecules, but even here it is not clear yet whether highly optimized conventional SCF-GTF expansion calculations are not just as rapid or even faster.[126] However, it is hoped that the $X\alpha$ method, young as it is, will still be improved and studied with respect to the reliability of computed molecular properties other than orbital energies, and that, if necessary, some of the approximations made can be compensated.

Acknowledgment

It is my pleasure to thank Ch. Eaker and R. Saxon for carefully reading the manuscript, and for their help with the bibliography. Thanks is also due to many of the active researchers in the area of large molecular calculations who have assisted by pointing out to me many of the calculations reported in the literature. I apologize for not being able to cite all these references; however, I found the task of compiling a comprehensive, annotated bibliography of computed molecular wavefunctions too monumental, although I consider it extremely important. Gratefully acknowledged is the support of the National Science Foundation, making this work possible.

References

1. M. Born and J. R. Oppenheimer, *Ann. Physik.*, **84**, 457 (1927); a modern treatment is given by M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, London, 1954. See also J. C. Slater, *Quantum Theory of Molecules and Solids*, 1963, p. 9, F. L. Pilar, *Elementary Quantum Chemistry*, 1968, p. 414; A. S. Davydov, *Quantum Mechanics*, p. 472.
2. C. E. Eckart, *Phys. Rev.*, **36**, 878 (1930); A. S. Davydov, *Quantum Mechanics*, 1965, p. 179.
3. E. A. Hylleraas and B. Undheim, *Z. Physik*, **65**, 759 (1930).
4. J. K. L. MacDonald, *Phys. Rev.*, **43**, 830 (1933).
5. D. R. Hartree, *Proc. Cambridge Phil. Soc.*, **24**, 89 (1928).
6. V. Fock, *Z. Physik*, **61**, 126 (1930).
7. J. C. Slater, *Phys. Rev.*, **35**, 210 (1930).

8. C. C. J. Roothaan, *Rev. Mod. Phys.*, **23,** 69 (1951).
9. G. G. Hall, *Proc. Roy. Soc.* (*London Ser. A*, **205,** 541 (1951).
10. R. Lefebvre, *J. Chim. Phys.*, **54,** 168 (1957).
11. C. C. J. Roothaan, *Rev. Mod. Phys.*, **32,** 179 (1960).
12. S. Huzinaga, *Phys. Rev.*, **122,** 131 (1961).
13. F. W. Birss and S. Fraga, *J. Chem. Phys.*, **38,** 2252 (1963); **40,** 3203, 3207, 3212 (1964).
14. C. C. J. Roothaan and P. S. Bagus, *Methods Comp. Phys.*, **2,** 47 (1963).
15. R. McWeeny, *Proc. Roy. Soc. London Ser. A*, **235,** 496 (1956); *Phys. Rev.*, **126,** 1028 (1962).
16. W. J. Hunt, T. H. Dunning, and W. A. Goddard, *Chem. Phys. Lett.*, **3,** 609 (1969).
17. P. O. Lowdin, *Advan. Chem. Phys.*, **2,** 207 (1959).
18. H. F. Schaefer, *The Electronic Structure of Atoms and Molecules*, Addison-Wesley, Reading, Mass., 1972.
19. P. O. Lowdin, *Phys. Rev.*, **97,** 1474 (1955).
20. E. R. Davidson, *Rev. Mod. Phys.*, **44,** 451 (1972).
21. R. McWeeny, *Rev. Mod. Phys.*, **32,** 335 (1960). R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics*, Academic Press, New York, 1969.
22. J. Frenkel, *Wave Mechanics*, Vol. 2, *Advanced General Theory*, Clarendon Press, Oxford, 1934, p. 460.
23. G. Das and A. C. Wahl, *J. Chem. Phys.*, **44,** 87 (1966); **56,** 1769 (1972).
24. J. Hinze, *J. Chem. Phys.*, **59,** in press (1973).
25. T. L. Gilbert, *Phys. Rev.*, **A6,** 580 (1972).
26. J. C. Slater, *Phys. Rev.*, **36,** 57 (1930).
27. A. C. Wahl, P. E. Cade, and C. C. J. Roothaan, *J. Chem. Phys.*, **41,** 2578 (1964).
28. L. Salmon and K. Ruedenberg, *Intern. J. Quantum Chem.*, **6,** 353 (1972). E. O. Steinborn and K. Ruedenberg, *Intern. J. Quantum Chem.*, **6,** 413 (1972).
29. S. F. Boys, *Proc. Roy. Soc.* (*London*) *Ser. A*, **200,** 542 (1950); **201,** 125 (1950).
30. H. Preuss, *Z. Naturforsch.*, **11,** 823 (1956).
31. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1965. J. H. Wilkinson and C. Reinsch, *Handbook for Automatic Computation*, Vol. II, *Linear Algebra*, Springer, New York, 1971.
32. R. S. Martin and J. H. Wilkinson, *Num. Math.*, **12,** 377 (1968); A. Dubrelle, *Num. Math.*, **15,** 450 (1970).
33. R. K. Nesbet, *J. Chem. Phys.*, **43,** 311 (1965).
34. C. F. Bender, R. P. Hosteny, A. Pipano, and I. Shavitt, *J. Com. Phys.*, **11,** 90 (1973).
35. J. D. Weeks, A. Hazi, and S. A. Rice, *Advan. Chem. Phys.*, **16,** 283 (1969).
36. W. H. Fink, *J. Chem. Phys.*, **57,** 1822 (1972).
37. M. E. Schwartz and J. D. Switalski, *J. Chem. Phys.*, **57,** 4125 (1972); L. Szasz and G. McGinn, *J. Chem. Phys.*, **56,** 1019 (1972).
38. C. C. J. Roothaan, L. M. Sachs, and A. W. Weiss, *Rev. Mod. Phys.*, **32,** 186 (1960).
39. E. Merzbacher, *Quantum Mechanics*, Wiley, New York, 1961, p. 188; see also Ref. 57.
40. A. D. McLean, *Proceedings of the Conference on Potential Energy Surface in Chemistry*, Publication RA 18, IBM Research Laboratory, San Jose, Cal., 1971.
41. F. E. Harris and H. H. Michels, *Advan. Chem. Phys.*, **13,** 205 (1967). K. G. Kay and H. J. Silverstone, *J. Chem. Phys.*, **51,** 956, 4287 (1969); **53,** 4269 (1970).

42. R. M. Stevens, *J. Chem. Phys.*, **52,** 1397 (1970). A. C. Wahl and R. H. Land, *J. Chem. Phys.*, **50,** 4725 (1969).
43. F. E. Harris, *Rev. Mod. Phys.*, **35,** 558 (1963); I. Shavitt, *Methods Comp. Phys.*, **2,** 1 (1963).
44. POLYATOM: I. G. Csizmadia, M. C. Harrison, J. W. Moskowitz, and B. T. Sutcliffe, *Theoret. Chim. Acta*, **6,** 191 (1966).
IBMOL: E. Clementi and D. R. Davis, *Intern. Comp. Phys.*, **2,** 223 (1967).
MOSES: L. M. Sachs and M. Geller, *Intern. J. Quantum Chem.*, **1S,** 445 (1967).
45. J. L. Whitten, *J. Chem. Phys.*, **44,** 359 (1966).
46. S. Huzinaga and C. Arnau, *J. Chem. Phys.*, **53,** 451 (1970). H. Basch, M. B. Robin, and N. A. Kuebler, *J. Chem. Phys.*, **47,** 201 (1967).
47. R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, **54,** 724 (1972). W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56,** 2257 (1972).
48. R. E. Christoffersen, L. L. Shipman, and G. M. Maggiora, *Intern. J. Quantum Chem.*, **5S,** 143 (1971).
49. E. Clementi, H. Kistenmacher, and H. Popkie, *J. Chem. Phys.*, **58,** 4699 (1973).
50. K. Ruedenberg, *Rev. Mod. Phys.*, **34,** 326 (1962).
51. P. S. Bagus, T. L. Gilbert, and C. C. J. Roothaan, *J. Chem. Phys.*, **56,** 5195 (1972).
52. E. Clementi, "Tables of Atomic Functions," *IBM J. Res. Develop. Suppl.*, **9,** 2 (1965). E. Clementi, *J. Chem. Phys.*, **40,** 1944 (1964). E. Clementi, R. Matcha, and A. Veillard, *J. Chem. Phys.*, **47,** 1865 (1967).
53. T. H. Dunning, *J. Chem. Phys.*, **53,** 2823 (1970).
54. P. C. Hariharan and J. A. Pople, *Theoret. Chim. Acta*, **28,** 213 (1973). P. S. Bagus, B. Liu, A. D. McLean, and M. Yoshimine, *Computational Methods for Large Molecules and Localized States in Solids*, Plenum Press, New York, 1973.
55. R. M. Stevens, *J. Chem. Phys.*, **55,** 1725 (1971).
56. B. J. Ransil, *J. Chem. Phys.*, **34,** 2109 (1961).
57. N. C. Handy, M. T. Marron, and H. J. Silverstone, *Phys. Rev.*, **180,** 45 (1969).
58. H. Preuss, *Intern. J. Quantum Chem.*, **2,** 651 (1968).
59. R. A. Rouse and A. A. Frost, *J. Chem. Phys.*, **50,** 1705 (1969). A. A. Frost, *J. Phys. Chem.*, **72,** 1289 (1968).
60. L. L. Shipman and R. E. Christofferson, *Chem. Phys. Lett.*, **15,** 469 (1972).
61. H. P. Kelley, *Advan. Chem. Phys.*, **14,** 129 (1969).
62. L. Hedin and S. Lundquist, *Solid State Phys.*, **23,** 1 (1969).
63. L. Brillouin, *Actualites Sci. Ind.*, **71** (1933); **159** (1934).
64. D. H. Kobe, *Am. J. Phys.*, **34,** 1150 (1966).
65. B. Levy and G. Berthier, *Intern. J. Quantum Chem.*, **2,** 307 (1968); **3,** 247 (1969).
66. C. Moller and M. S. Plesset, *Phys. Rev.*, **46,** 618 (1934). G. G. Hall, *Phil. Mag.*, **6,** 249 (1961). M. Cohen and A. Dalgarno, *Proc. Phys. Soc.* London, **77,** 748 (1961).
67. C. F. Bender and E. R. Davidson, *J. Chem. Phys.*, **49,** 4222 (1968).
68. P. Bertoncini and A. C. Wahl, *Phys. Rev. Lett.*, **25,** 991 (1970). P. J. Bertoncini, G. Das, and A. C. Wahl, *J. Chem. Phys.*, **52,** 5112 (1970). A. C. Wahl, *J. Chem. Phys.*, **41,** 2600 (1964).
69. L. C. Snyder and H. Basch, *J. Am. Chem. Soc.*, **91,** 2189 (1969). W. J. Hehre, L. Radom and J. A. Pople, *J. Am. Chem. Soc.*, **92,** 4796 (1970).
70. T. A. Koopmans, *Physica*, **1,** 104 (1933).
71. P. S. Bagus and H. F. Schaefer, *J. Chem. Phys.*, **55,** 1474 (1971).
72. H. J. Silverstone and M. L. Yin, *J. Chem. Phys.*, **49,** 2026 (1968).
73. S. Huzinaga and C. Arnau, *J. Chem. Phys.*, **54,** 1948 (1971).

73. S. Huzinaga and C. Arnau, *J. Chem. Phys.*, **54,** 1948 (1971). W. J. Hunt and W. A. Goddard, *Chem. Phys. Lett.*, **3,** 414 (1969).
74. E. R. Davidson, *J. Chem. Phys.*, **57,** 1999 (1972). S. T. Elbert, S. R. Langhoff, and E. R. Davidson, *J. Chem. Phys.*, **57,** 2005 (1972).
75. H. Basch and V. McKoy, *J. Chem. Phys.*, **53,** 1628 (1970). T. H. Dunning, W. J. Hunt, and W. A. Goddard, *Chem. Phys. Lett.*, **4,** 147 (1969).
76. T. Shibuya and V. McKoy, *J. Chem. Phys.*, **54,** 1738 (1971). J. Rose, T. Shibuya, and V. McKoy, *J. Chem. Phys.*, **58,** 74 (1973).
77. D. J. Rowe, *Rev. Mod. Phys.*, **40,** 153 (1968). T. Shibuya and V. McKoy, *Phys. Rev.*, **A2,** 2208 (1970). T. Shibuya, J. Rose and V. McKoy, *J. Chem. Phys.*, **58,** 500 (1973).
78. R. J. Buenker, S. D. Peyerimhoff, and W. E. Kammer, *J. Chem. Phys.*, **55,** 814 (1971).
79. L. L. Shipman and R. E. Christoffersen, *Proc. Natl. Acad. Sci. U.S.*, **69,** 3301 (1972). D. W. Genson and R. E. Christoffersen, *J. Am. Chem. Soc.*, **95,** 362 (1973).
80. E. Clementi, *Proc. Natl. Acad. Sci. U.S.*, **69,** 2942 (1972).
81. E. Clementi, J. Mehl, and W. von Niessen, *J. Chem. Phys.*, **54,** 508 (1971).
82. L. R. Kahn and W. A. Goddard, *J. Chem. Phys.*, **56,** 2685 (1972). C. F. Melius and W. A. Goddard, *J. Chem. Phys.*, **56,** 3348 (1972).
83. R. E. Christoffersen, *Advan. Quantum Chem.*, **6,** 333 (1972).
84. W. A. Lathan, L. Radom, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, **95,** 699 (1973).
85. L. Radom, W. A. Lathan, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, **95,** 693 (1973).
86. W. P. Kraemer and G. H. F. Diercksen, *Theoret. Chim. Acta*, **27,** 265 (1972).
87. G. H. F. Diercksen, W. P. Kraemer, and W. von Niessen, *Theoret. Chim. Acta*, **28,** 67 (1972).
88. H. Kistenmacher, E. Clementi, and H. Popkie, *J. Chem. Phys.*, **58,** 1689 (1973).
89. S. Shih, R. J. Buenker, S. D. Peyerimhoff, and C. J. Michejda, *J. Am. Chem. Soc.*, **94,** 7621 (1972).
90. E. Clementi, J. Mehl, and W. von Niessen, *J. Chem. Phys.*, **54,** 508 (1971).
91. J. M. Lehn and G. Wipff, *Theoret. Chim. Acta*, **28,** 223 (1973).
92. J. Demuynek and A. Veillard, *Theoret. Chim. Acta*, **28,** 241 (1973).
93. R. E. Christoffersen, *J. Am. Chem. Soc.*, **93,** 4104 (1971).
94. S. Shih, R. J. Buenker, and S. D. Peyerimhoff, *Chem. Phys. Lett.*, **16,** 244 (1972).
95. R. Ditchfield, J. Del Bene, and J. A. Pople, *J. Am. Chem. Soc.*, **94,** 4806 (1972).
96. M. Krauss, *Compendium of ab initio Calculations of Molecular Energies and Properties*, NBS Technical Notes 438, 1967. W. G. Richards, T. E. H. Walker, and R. K. Hinkley, *A Bibliography of ab initio Molecular Wave Functions*, Clarendon Press, Oxford, 1971. L. C. Snyder and H. Basch, *Molecular Wave Functions and Properties*, Wiley, New York, 1972.
97. L. Hedin and S. Lundquist, *Solid State Phys.*, **23,** 1 (1969). G. Verhaegen, *J. Chem. Phys.*, **49,** 4696 (1968).
98. J. A. Pople, D. P. Santry, and G. A. Segal, *J. Chem. Phys.*, **43,** S129 (1965).
99. J. A. Pople, D. L. Beveridge, and P. A. Dobosh, *J. Chem. Phys.*, **47,** 2026 (1967).
100. N. C. Baird and M. J. S. Dewar, *J. Chem. Phys.*, **50,** 1262 (1969).
101. M. J. S. Dewar and G. Klopman, *J. Am. Chem. Soc.*, **89,** 3089 (1967).
102. G. Klopman and B. O'Leary, *Fortschr. Chem. Forsch.*, **15** (4) 445 A. Pullman, *Fortschr. Chem. Forsch.*, **31,** 45 (1970).

103. J. A. Pople and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.
104. M. J. S. Dewar, *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill, New York, 1969.
105. J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **43,** 5129 (1965); and Ref. 103.
106. H. Fischer and H. Kollmar, *Theoret. Chim. Acta*, **13,** 213 (1969).
107. N. Bodar, M. J. S. Dewar, A. Harget, and E. Haselbach, *J. Am. Chem. Soc.*, **92,** 3854 (1970).
108. J. Del Bene and H. H. Jaffé, *J. Chem. Phys.*, **48,** 1807, 4050 (1968); **49,** 1221 (1968).
109. Several programs are available through the Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington, Indiana.
110. R. Hoffmann, *J. Chem. Phys.*, **39,** 1397 (1963).
111. A. Streitwieser, *Molecular Orbital Theory for Organic Chemists*, Wiley, New York, 1961.
112. R. B. Woodward and R. Hoffmann, *The Conservation of Orbital Symmetry*, Academic Press, New York, 1969.
113. R. S. Mulliken, *J. Chim. Phys.*, **46,** 497, 675 (1949).
114. K. Ruedenberg, *J. Chem. Phys.*, **19,** 1433 (1951) and extensions; G. Berthier, P. Millie, and A. Veillard, *J. Chim. Phys.*, **62,** 8 (1965).
115. B. J. Nicholson, *Adv. Chem. Phys.*, **18,** 249 (1970).
M. D. Newton, *J. Chem. Phys.*, **51,** 3917 (1969).
116. M. D. Newton, N. S. Ostlund, and J. A. Pople, *J. Chem. Phys.*, **49,** 5192 (1968).
117. F. P. Billingsley and J. E. Bloor, *J. Chem. Phys.*, **55,** 5178 (1971).
118. K. H. Johnson, J. G. Norman, and J. W. D. Connolly, *Computational Methods for Large Molecules and Localized States in Solids*, Plenum Press, New York, 1973, p. 161.
119. J. C. Slater and K. H. Johnson, *Phys. Rev.*, **B5,** 844 (1972). K. H. Johnson and F. C. Smith, *Phys. Rev.*, **B5,** 831 (1972).
120. J. Korringa, *Physica*, **13,** 392 (1947). W. Kohn and N. Rostoker, *Phys. Rev.*, **94,** 1111 (1954).
121. J. C. Slater, *Advances in Quantum Chemistry*, Vol. 6, Academic Press, 1973.
122. J. C. Slater, T. M. Wilson, and J. H. Wood, *Phys. Rev.*, **179,** 28 (1969); J. C. Slater and J. H. Wood, *Intern. J. Quantum Chem.* **4S,** 3 (1971).
123. K. H. Johnson and F. C. Smith, *Intern. J. Quantum Chem.* **5S,** 429 (1971).
124. J. G. Norman, K. H. Johnson, and F. A. Cotton, unpublished; see Ref. 118.
125. J. W. D. Connolly and J. R. Sabin, *J. Chem. Phys.* **56,** 5529 (1972)
126. The calculation of the barrier of rotation of C_2H_6 using the X_α method was reported by U. Wahlgren and K. H. Johnson, *J. Chem. Phys.*, **56,** 3715 (1972). A comparable calculation using GTE's in a conventional SCF calculation see [Ref. 49] was shown to be two to three times faster.

THE TWO-COMPONENT BÉNARD PROBLEM

R. S. SCHECHTER, M. G. VELARDE*, AND J. K. PLATTEN†

Center for Statistical Mechanics and Thermodynamics
Austin, Texas

CONTENTS

I. Concentration Gradients and Hydrodynamic Stability 266
II. Formulation of the Problem 268
A. The Conservation Equations 268
B. The Thermohaline Problem 269
C. Thermal Diffusion 272
D. Chemical Reactions 272
III. Thermohaline Convection: Linear Analysis 272
A. "Similar" Boundary Conditions 272
B. Free Boundaries with Specified Solute Concentrations and Temperatures 275
C. General Boundary Conditions 279
D. Experimental Observations 279
IV. Convective Motion and Thermal Diffusion: Linear Theory . . . 282
A. The Temperature Coupled Mass Balance 282
B. Free Boundaries 284
C. Impermeable Conducting Boundaries 285
D. An Approximation 289
E. Experimental Results 289
V. Chemical Instabilities 293
VI. Global Stability: Energy Methods 295
VII. Supercritical States 296
A. A Classification of Problems 296
B. Stabilizing Solute Gradients 297
C. Destabilizing Solute Gradients 298
Acknowledgment 299
References 299

* Present address: Departamento de Fisica, Universidad autonoma de Madrid, Canto Blanco, Madrid, Spain.

† Present address: Department of Thermodynamics, University of Mons, 7000 Mons, Belgium.

I. CONCENTRATION GRADIENTS AND HYDRODYNAMIC STABILITY

Our interest in the two-component Bénard problem was stimulated by a suggestion of Professor I. Prigogine that the Soret coefficient should be related to the instability of a binary fluid heated from below. The idea was that the critical temperature gradient, that is, the temperature gradient required to initiate convection currents, should depend on the extent of the thermal diffusion. As will be seen, this idea has now been exploited to yield a means of measuring the Soret coefficient. Furthermore, there has been considerable related work which has indicated the profound influence of small concentrations on the instability of a heated fluid layer. This has far-reaching implications ranging from the operation of solar ponds[69] to the thermal microstructure of the ocean,[10] and perhaps even to convection in stars.[49] This review is intended to unify a number of apparently different and often surprising phenomena. For example, when a layer of warm salty water is carefully poured on top of the surface of a cold, more dense layer of salt free water, the interface between the two layers becomes turbulent. A structure resembling a series of fingers then spontaneously develops and grows in length.[47] Convective mixing between the two phases is observed even though the overall density gradient favors stability with the more dense fluid being on the bottom!

If a layer of water is heated from below and/or cooled from above, a stationary (time independent) convective motion ensues when the applied temperature gradient exceeds a certain critical value. If a small amount of isopropanol (10 wt %) is added the observed critical temperature difference is increased by 40%. Furthermore, the convection currents exhibit time oscillations.[32,34] A 3 wt % saltwater solution having an average temperature of 6°C exhibits the same behavior. If, on the other hand, the average temperature of the saltwater layer is increased to 20°C, the critical temperature difference required to yield increased heat transfer across the fluid layer is about the same as that for pure water and the salt appears to have no effect on the heat flux. (We shall find, however, that observing the heat flux is not a sensitive technique for detecting the onset of convection.) In this case the convective motion is nonoscillatory.

If an aqueous solution of KI is heated from above and cooled from below, convection currents can be induced by exceeding a certain critical temperature gradient. Convection currents cannot, on the other hand, be nucleated by heating an aqueous solution of KCl from above regardless of the imposed temperature gradient.

These strange phenomena and related topics such as the presence of nonsteady convection currents in melts supporting the growth of single

crystals are all related to convective instabilities triggered in many cases by exceedingly small gradients in concentration. Concentration fluctuations in liquids are far more dangerous than thermal, velocity, density, or electrical potential fluctuations because of their relatively long lifetime. The effect of this tendency toward the long persistence of concentration fluctuations can be visualized in the following way. Suppose that a small parcel of fluid takes a small vertical displacement in a liquid layer in which the temperature and the salinity both increase with elevation. Assume that the parcel is initially more dense than the surrounding fluid because it is cooler. The temperature of the fluid parcel will, however, become rapidly equalized with that of the surrounding fluid and the parcel will then be less dense because of its lower salinity. It will again rise, therefore, and the process will be repeated. Convection currents, called salt fingers, will result. Thus this convective instability, which is triggered even though the overall density gradient is not adverse, owes its existence to the difference between the relaxation times for thermal and concentration fluctuations.

A little thought will reveal the ubiquitous nature of this particular mechanism for producing instability. It can, and does, exist in three-component diffusion because the relaxation times (diffusion coefficients) can differ among the diffusing components. Fine suspensions would also be prime candidates because the Brownian diffusion coefficients are much smaller than thermal conductivities. In fact any system in which the fluid density depends on two variables that diffuse at differing rates is liable to exhibit spontaneous convection and often at the most unexpected times.

The purpose of this chapter is to collect and review the literature related to the processes described above so that they can be discussed quantitatively. Most of the emphasis will be placed on linear theory which in reality is a study of the fate of small perturbations. To complete the analysis we should also consider the stability with respect to arbitrary perturbations, but the mathematical tools available to treat such questions are still relatively primitive. A few studies of the difficult, nonlinear two-component Bénard problem have been reported and are summarized in this article. Questions concerning the possible existence of subcritical, finite amplitude instabilities are generally answered using Liapunov methods which provide sufficient but not necessary conditions for stability. The energy methods fall into this category. They define regions in which stability is certain, but do not help in the precise demarcation of stable from unstable regions. Glansdorff and Prigogine[12] have made extensive use of Liapunov functions in their comprehensive treatment of stability.

A second, and perhaps more complex, task is that of determining the particular state to which a system will evolve from an initially unstable state. Perturbation techniques have usually been used to help resolve this

problem, but Koschmieder[19] questions the success of these techniques in this volume. He points out that the experimental results relating to the classical Bénard problem in the supercritical region are not predicted by the present theories and further work in this important area is definitely required.

II. FORMULATION OF THE PROBLEM

A. The Conservation Equations

The system to be investigated is a transversely infinite, horizontal fluid layer of thickness d, initially in mechanical equilibrium but subjected to gradients of both temperature and concentration. The Boussinesq equations of motion (see Chandrasekhar[9]) are composed of the following equations expressing the conservation of momentum,

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho_m} \nabla p + \mathbf{g} \frac{\rho}{\rho_m} + \nu \nabla^2 \mathbf{v} \tag{2.1}$$

conservation of mass,

$$\nabla \cdot \mathbf{v} = 0 \tag{2.2}$$

conservation of energy,

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = -\frac{1}{\rho_m C_p} \nabla \cdot \mathbf{q} + \frac{1}{\rho_m C_p} G \tag{2.3}$$

and the conservation of individual species,

$$\frac{\partial S_i}{\partial t} + \mathbf{v} \cdot \nabla S_i = -\frac{1}{\rho_m} \nabla \cdot \mathbf{j}_i + \frac{1}{\rho_m} J_i \tag{2.4}$$

The density of the fluid has been taken to be the mean density of the layer, ρ_m, except in the gravitational term of (2.1). Furthermore, all the physical properties such as the kinematic viscosity, ν, and the heat capacity, C_p, will be assumed constant throughout the analysis. The internal generation of heat, G, and the mass, J_i, must be included in any treatment of reacting systems.

The linearized equation of state

$$\rho = \rho_m[1 - \alpha(T - T_m) + \sum \beta_i(S_i - S_i^m)] \tag{2.5}$$

where

$$\alpha = -\frac{1}{\rho} \frac{\partial \rho}{\partial T}$$

and

$$\beta_i = \frac{1}{\rho} \frac{\partial \rho}{\partial S_i}$$

is used in all subsequent calculations. The mean density, ρ_m, is evaluated at the mean temperature, T_m, and average mass fraction, S_i^m.

Since we are interested in the stability of a quiescent layer of fluid oriented horizontally and subjected to linear gradients of both temperature and solute concentration, the stationary state to be tested for stability is, in the absence of generation of heat or mass, characterized by the following equations:

$$\mathbf{v}^s = 0 \tag{2.6}$$

$$T^s = T_0 - \Delta T \frac{z}{d} \tag{2.7}$$

$$S_i^s = S_i^0 - \Delta S_i \frac{z}{d} \tag{2.8}$$

where T_0 and S_i^0 are the temperature and the mass fractions evaluated at the lower surface. ΔT and ΔS_i are differences taken between the lower and upper surface so that $\Delta T > 0$ if the temperature of the lower surface is warmer than the upper one. The steady-state pressure can be determined using (2.1) and (2.5) to be given by

$$\frac{1}{\rho_m}\frac{dp^s}{dz} = -g[1 - \alpha(T^s - T_m) + \sum \beta_i(S_i - S_i^m)] \tag{2.9}$$

To study the regression of small fluctuations about this steady state, we can consider the linearized conservation equations as follows

$$\frac{\partial\, \delta\mathbf{v}}{\partial t} = -\frac{1}{\rho_m}\nabla\, \delta p + \mathbf{g}\frac{\delta\rho}{\rho_m} + \nu\nabla^2\, \delta\mathbf{v} \tag{2.10}$$

$$\nabla \cdot \delta\mathbf{v} = 0 \tag{2.11}$$

$$\frac{\partial\, \delta T}{\partial t} - w\frac{\Delta T}{d} = -\frac{1}{\rho_m C_p}\nabla \cdot \delta\mathbf{q} + \frac{1}{\rho_m C_p}\delta G \tag{2.12}$$

and

$$\frac{\partial\, \delta S_i}{\partial t} - w\frac{\Delta S_i}{d} = -\frac{1}{\rho_m}\nabla \cdot \delta\mathbf{j}_i + \frac{1}{\rho_m}\delta J_i \tag{2.13}$$

where $\delta(\ )$ represents a small fluctuation about the steady-state and $\delta\mathbf{v} = (u, v, w)$.

B. The Thermohaline Problem

If the gradients of mass and temperature are independent and if there are no chemical reactions then

$$\delta G = \delta J_i = 0$$

and

$$\mathbf{j}_i = -\rho_m D \nabla S_i$$

and

$$\frac{1}{\rho_m C_p} \mathbf{q} = -\kappa \nabla T$$

where κ is the thermal diffusivity and D is the molecular diffusivity. Convection currents driven by small differences in solute concentration have been termed thermohaline convection by Stern[50] because of the possible importance of these currents in the vertical mixing of the sea. This convection has also been called salt fingers, because the cell structure of the instability appears to tend to a pattern resembling long fingers.

Assuming that the horizontal layer is infinite in extent, then perturbations of the form

$$\delta T = |\Delta T|\, \Theta(Z) \exp\left[i(k_X X + k_Y Y) - \sigma\tau\right]$$
$$\delta S_i = \Gamma_i(Z) \exp\left[i(k_X X + k_Y Y) - \sigma\tau\right]$$
$$\delta W = \frac{\nu}{d} w(Z) \exp\left[i(k_X X + k_Y Y) - \sigma\tau\right]$$

and similar forms for the other dependent variables are conveniently studied. Here Θ, Γ_i, and W are dimensionless amplitudes which can be shown to be related by the following differential equations:

$$0 = -\left(\frac{d^2}{dZ^2} - k^2\right)\left(\frac{d^2}{dZ^2} - k^2 + \sigma\right) W + \frac{k^2}{Pr} R\Theta - \sum R_s^{(i)} \frac{k^2}{\Delta S_i} \frac{\Gamma_i}{Sc_i} \tag{2.14}$$

$$0 = \sigma\Theta + \frac{\Delta T}{|\Delta T|} W + \frac{1}{Pr}\left(\frac{d^2}{dZ^2} - k^2\right)\Theta \tag{2.15}$$

$$0 = \sigma\Gamma_i + \Delta S_i W + \frac{1}{Sc_i}\left(\frac{d^2}{dZ^2} - k^2\right)\Gamma_i \tag{2.16}$$

This coupled set of homogeneous equations follows directly from (2.10), (2.11), (2.12), and (2.13). The perturbations are in effect Fourier components of an arbitrary disturbance in the horizontal plane. The wave number $k = [= (k_x^2 + k_y^2)^{1/2}]$ identifies the particular Fourier component and, because the system is infinite horizontally, all real positive values of k are admissible. The dimensionless damping factor, σ, can be complex but if its real part is negative, then the perturbation will grow exponentially in time $[\tau = (\nu/d^2)t]$ and the convection free state is unstable. States for which the real part of σ vanishes are called marginal states. If the imaginary part of σ is nonzero at a point on the neutral stability curve, then the marginal state is an oscillatory one and any perturbation of temperature, velocity, or concentration about this state will result in sustained temporal oscillations.

The state of the system is determined by the dimensionless parameters R, Pr, Sc_i, and $R_s^{(i)}$. R is the Rayleigh number $(= gd^3 \alpha \Delta T/\kappa\nu)$ and is a measure of the buoyancy force created by the temperature gradient relative to the viscous forces. It can be either positive or negative depending on the sign of ΔT. For the multicomponent problem each of $N-1$ (N is the total number of components) species can also contribute to the buoyancy and this contribution is measured by the solute Rayleigh number, $R_s^{(i)}$ $(= gd^3\beta_i \Delta S_i/D\nu)$. The physical properties are given in terms of a Prandtl number, Pr $(= \nu/\kappa)$, and $N-1$ Schmidt numbers, Sc_i $(= \nu/D_i)$. It should be noted that the dimensionless time and distances are related to their dimensional counterparts as follows: $\tau = (\nu/d^2)t, X = x/d$, $Y = y/d$, and $Z = z/d$.

Equations (2.14), (2.15), and (2.16) are to be solved subject to appropriate boundary conditions. For the velocity there are two models in general use: free or rigid boundaries. At a free boundary

$$W = \frac{d^2W}{dZ^2} = 0 \tag{2.17}$$

and on a rigid boundary

$$W = \frac{dW}{dZ} = 0 \tag{2.18}$$

The free boundary conditions arises because there can be neither normal flow nor shear stress at a free boundary and the rigid boundary condition expresses the necessity for all components of velocity to vanish at the boundary.

The boundary conditions imposed upon the temperature fluctuations can be expressed as follows:

$$\frac{d\Theta}{dZ} + L\Theta = 0 \tag{2.19}$$

where L_u and L_l refer to the upper and lower boundaries, respectively. There are two special cases of particular interest. $L \to \infty$ implies $\Theta \to 0$. The temperature fluctuations vanish and the temperature is therefore fixed. This is modeled in the laboratory by employing a highly conductive boundary. At the other extreme, $L \to 0$ yields $d\Theta/dz = 0$, which is modeled by a bounding surface of poor conductivity.

Similarly the boundary conditions to be imposed on the perturbations of mass fraction can be written as follows:

$$\frac{d\Gamma_i}{dZ} + M^{(i)}\Gamma_i = 0 \tag{2.20}$$

With $M^{(i)} = M_u^{(i)}$ at $Z = 1$ and $M^{(i)} = M_l^{(i)}$ at $Z = 0$.

Various solutions of the thermohaline problem are to be presented in Section III.

C. Thermal Diffusion

Our theme is that small concentration gradients disproportionately influence hydrodynamic stability relative to their contribution to the buoyancy of a fluid. Section III is devoted to problems for which the concentration and temperature gradients are independent. The results obtained in Section III also apply when the concentration gradients arise because of thermal diffusion, but because of the potential importance of thermal diffusion we have elected to consider it in a separate section.

D. Chemical Reactions

Section V also deals with concentration gradients which may be coupled to the temperature. Chemical reactions in multicomponent systems may yield a variety of products at a rate that depends on the temperature, or the local equilibrium composition of a multicomponent mixture may very well depend on the local temperature.

Either a temperature-dependent reaction rate or equilibrium will give rise to concentration gradients that depend on the imposed temperature. Thus the effect of chemical reactions is similar to thermal diffusion. The two mechanisms are treated in separate sections, however, because they are formulated quite differently. The separation arising from thermal diffusion is quantified by adding an appropriate correction term to the mass flux relationship (see Section IV) whereas chemical reactions change the generation terms G and J_i of (2.3) and (2.4). Because of this mathematical difference, the two mechanisms are treated separately.

III. THERMOHALINE CONVECTION: LINEAR ANALYSIS

A. "Similar" Boundary Conditions

We can make some rather general remarks concerning linear stability if the boundary conditions imposed on the temperature fluctuations are "similar" to those imposed on the concentration fluctuations, that is, if

$$M_l^{(i)} = L_l \qquad \text{for all } i$$

and

$$M_u^{(i)} = L_u \qquad \text{for all } i$$

Then it can be shown that the conditions for stationary stability reduce exactly to those of the pure component case. The Rayleigh number for the pure component case, however, must be replaced by an effective one defined as follows:

$$R_{\text{eff}} = R - \sum R_s^{(i)} \tag{3.1}$$

The critical effective Rayleigh number is, therefore, numerically equivalent to the critical Rayleigh number for the single-component system and the critical wave number is the same for all combinations of R, $R_s^{(i)}, \ldots$. Thus the flow patterns induced by thermal effects are identical to those produced by solute gradients or by a combination of thermal and solute gradients.

Furthermore, the hypersurface defining neutral stability in R, $R_s^{(1)}$ $R_s^{(2)}, \ldots$ space is therefore a hyperplane.

The proof that thermal and solutal effects are similar if the boundary conditions are similar can be demonstrated by defining an effective fluctuation

$$R_{\text{eff}}\psi = \frac{\Delta T}{|\Delta T|}\frac{R}{Pr}\Theta - \sum R_s^{(i)}\frac{1}{Sc_i}\frac{\Gamma_i}{\Delta S_i} \tag{3.2}$$

and upon setting $\sigma = 0$ (the condition for a stationary marginal state), (2.14), (2.15), and (2.16) can be shown to reduce to

$$-\left(\frac{d^2}{dZ^2} - k^2\right)^2 W + k^2 R_{\text{eff}}\psi = 0 \tag{3.3}$$

and

$$\left(\frac{d^2}{dZ^2} - k^2\right)\psi + W\frac{\Delta T}{|\Delta T|} = 0 \tag{3.4}$$

subject to the boundary conditions

$$\frac{d\psi}{dZ} + M_l\psi = 0 \qquad \text{at} \qquad Z = 0$$

and

$$\frac{d\psi}{dZ} + M_u\psi = 0 \qquad \text{at} \qquad Z = 1$$

Solutions to this problem are applicable to all combinations of thermal and solutal states provided the boundary conditions are similar. This relationship was first observed by Nield.[30] Listed in Table I are the available solutions for a variety of boundary conditions.

Several very interesting features of the thermohaline problem can now be noted. Consider the density profile of the stationary state being tested for stability. Equation (2.5) leads to the expression

$$\frac{\Delta\rho}{\rho_m} = \left(\frac{\nu\kappa}{gd^3}\right)\left\{-R + \sum R_s^{(i)}\frac{Pr}{Sc_i}\right\} \tag{3.5}$$

where $\Delta[\ ] = [\ ]_{\text{lower}} - [\ ]_{\text{upper}}$. If, therefore, $\Delta\rho > 0$, the system may be intuitively thought to be stable since the density *decreases* with increasing elevation. Such, however, is not the case. Consider a solvent

TABLE I

Critical Values of Rayleigh Number, $R_{\text{eff}}^{\text{crit}}$, and the Corresponding Wave Number k_c for Various Boundary Conditions. (The Results Apply to Stationary Stability, but Not to Overstable Cases[a])

Case	K_l	K_u	M_l	M_u	$R_{\text{eff}}^{\text{crit}}$	k_c	Reference
1	∞	∞	∞	∞	657.511	2.22	35
2	∞	∞	0	∞	384.693	1.76	30
3	∞	∞	0	0	120.0	0	30
4	0	∞	∞	∞	1100.657	2.68	36
5	0	∞	0	∞	816.748	2.21	48
6	0	∞	∞	0	669.001	2.09	48
7	0	∞	0	0	320	0	48
8	0	0	∞	∞	1707.765	3.12	36
9	0	0	0	∞	1295.781	2.55	48
10	0	0	0	0	720	0	48

[a] $K = 0$ rigid boundary; $K = \infty$, free boundary; $M = 0$ constant flux; $M = \infty$ constant temperature or solute concentration.

containing a single solute. The criterion for stationary instability is

$$R - R_s \geqslant R_{\text{eff}}^{\text{crit}}$$

when the boundary conditions are similar. Clearly, the condition for instability can be satisfied even when

$$\Delta\rho \sim -R + \frac{R_s P_r}{S_c} > 0$$

Indeed this is true if

$$\frac{R_s P_r}{S_c} \geqslant R \geqslant R_{\text{eff}}^{\text{crit}} + R_s \tag{3.6}$$

Thus instabilities are possible for negative values of R and R_s even when the density gradient is stable. It should be noted, however, that such an instability is only possible provided $S_c > P_r$. Thus this remarkable convective instability first discovered by Stommel et al.[53] owes its existence to the difference between the relaxation times for thermal and solutal fluctuations.

It should be noted that a related instability can be expected to exist even in the absence of a temperature gradient, provided the Schmidt numbers of a multicomponent system are different. Thus in the case of two solutes, the condition for the instability would be that the gradient of component 1 be a destabilizing one and $Sc_1 > Sc_2$. Since it is very unlikely that two solutes have precisely the same diffusion coefficient in any given

solvent, this instability is no doubt one that is quite common in natural processes as well as in many multicomponent systems touted as being convectionless because of the overall stabilizing density gradient. We see from our calculations here that creating a convectionless environment is a difficult task.

B. Free Boundaries with Specified Solute Concentrations and Temperatures

Studies of this problem have been reported by Stern,[50] Walin,[66] Nield,[30] and Baines and Gill.[2] It is particularly attractive to the theorist because the eigenfunctions are simple. Indeed, for a two-component mixture, the functions

$$[W, \Theta, \Gamma] = [A, B, C] \sin n\pi z$$

satisfy (2.14), (2.15), and (2.16) and the appropriate boundary conditions. A nontrivial solution is seen to exist if, and only if,

$$q^3 - q^2\left(1 + \frac{1}{Sc} + \frac{1}{Pr}\right) + q\left(\frac{1}{Pr} + \frac{1}{Sc} + \frac{1}{PrSc} - \frac{R'}{Pr} + \frac{R_s'}{Sc}\right) - \frac{1}{PrSc} + \frac{R'}{PrSc} - \frac{R_s'}{PrSc} = 0 \quad (3.7)$$

where

$$a^2 q = \sigma$$

$$R' = \frac{Rk^2}{a^6}$$

$$R_s' = \frac{R_s k^2}{a^6}$$

$$a^2 = n^2\pi^2 + k^2$$

Since Real$\{\sigma\} > 0$ implies that small fluctuations will decay, the neutral stability curve is found by setting

$$\text{Real}\{q\} = 0 \qquad \text{or} \qquad q = i\omega \text{ with } \omega \text{ real}$$

Equation (3.7) when separated into its real and imaginary parts yields

$$\omega^2\left(\frac{1}{Pr} + \frac{1}{Sc} + 1\right) - \frac{1}{PrSc}(1 - R + R_s') = 0 \quad (3.8)$$

and

$$\omega\left[-\omega^2 + \frac{1}{Pr} + \frac{1}{Sc} + \frac{1}{PrSc} - \frac{R'}{Pr} + \frac{R_s'}{Sc}\right] = 0 \quad (3.9)$$

Stationary or direct stability is found by setting $\omega = 0$. The equations then have the simple solution

$$R' - R_s' = 1 \tag{3.10}$$

or

$$R - R_s = \frac{(n^2\pi^2 + k^2)^3}{k^2} \tag{3.11}$$

The limits of linear stability are defined by seeking the wave numbers k and n which yield the minimum temperature and concentration gradients necessary to satisfy (3.11). For a finite depth, $n = 1$ is the smallest admissible value. Walin[66] has considered the results for small n (infinite depth), but we shall study finite depths. Clearly $n = 1$ and $k_c^2 = \pi^2/2$ minimize the right-hand side of (3.11) and therefore represent the critical values. Correspondingly,

$$R - R_s = \frac{27\pi^4}{4} = 657.511 = R_{\text{eff}} \tag{3.12}$$

This result is in fact predictable based on the arguments set forth in Section III.A, since the thermal and solutal boundary conditions are similar. The critical effective Rayleigh number corresponds to case 1 of Table I. The curve representing the marginal stationary state is shown in Fig. 1. Also shown is the curve representing neutral buoyancy, that is,

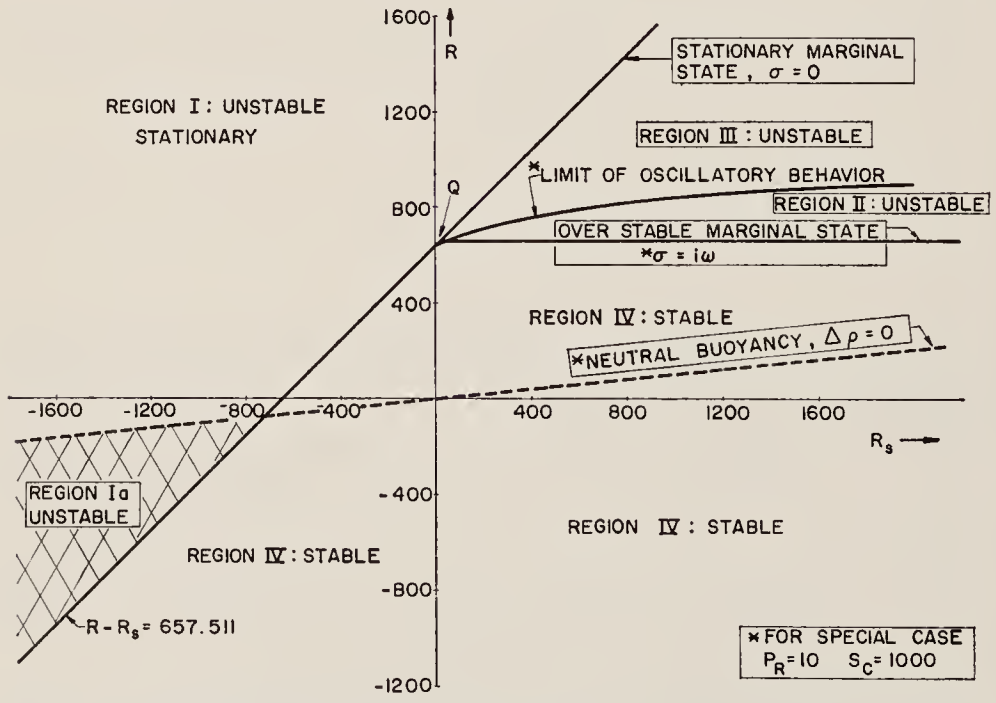


Fig. 1. The neutral stability curve for thermohaline convection. The layer is bounded by free surfaces having a constant temperature and composition.

$\Delta\rho = 0$. Points $[R, R_s]$ in Region Ia represent states for which a stationary convection driven by a solutal concentration gradient will exist even though the density gradient is not adverse. This is the remarkable thermohaline convection first identified by Stern.[50]

These neutral stability curves, obtained by setting $\omega = 0$, are included in the general arguments presented in Section III.A. To be complete, however, we must consider possible overstable states, that is, solutions to (3.8) and (3.9) for which $\omega \neq 0$. This solution is

$$\frac{RSc^2}{(Sc + Pr)(1 + Sc)} - \frac{R_s Pr^2}{(Sc + Pr)(1 + Pr)} = \frac{27\pi^4}{4} \tag{3.13}$$

with the additional stipulation

$$\omega^2 = \frac{1}{Pr} + \frac{1}{Sc} + \frac{1}{PrSc} - \frac{4}{27\pi^4}(R - R_s) \geqslant 0 \tag{3.14}$$

The line defined by (3.13) intersects the neutral stability curve for stationary marginal states at point Q (see Fig. 1) having the coordinates

$$\begin{aligned} (R_s)_Q &= \frac{27\pi^4}{4} \frac{(1 + Pr)}{(Sc - Pr)} \\ (R)_Q &= \frac{27\pi^4}{4} \frac{(1 + Sc)}{(Sc - Pr)} \end{aligned} \tag{3.15}$$

All points $[R, R_s]$ satisfying (3.13) together with $R > (R)_Q$ and $R_s > (R_s)_Q$ also satisfy the inequality (3.14) and are therefore solutions to (3.8) and (3.9). Points for which $R < (R)_Q$ and $R_s < (R_s)_Q$ do not satisfy inequality (3.14), and therefore overstability is possible only for $R_s \geqslant (R_s)_Q \geqslant 0$ for $Sc > Pr$. The solute gradient, therefore, must tend to be stabilizing before overstable states are possible. The line separating stable from overstable states terminates at point Q. Where they exist, overstable states can be realized at smaller Rayleigh numbers than can stationary states. This seems to have been observed first by Veronis.[64]

Baines and Gill[2] have noted that the region of possible overstable states is limited. Region II of Fig. 1 denotes that region in which the roots of (3.7) are complex with at least one root having a negative real part. To suggest that the linear theory used here can be extrapolated to define the limits of oscillatory motion, is conjecture. However, we note that Goroff[13] has measured heat convection for the closely analogous system of convection restrained by rotation, and found that the gradient of the curve of the Nusselt number versus Rayleigh number exhibits a sharp increase of the point where linear theory predicts that direct modes become possible. This suggests that finite amplitude oscillatory modes tend to be less efficient

than direct modes for transporting heat. Sani[38] has provided theoretical evidence which yields this same conclusion. It is possible, but not yet experimentally verified, that a similar change in both the solute and heat fluxes occurs near the transition from Region I to Region II.

For boundary conditions which are similar, the critical wave number will be a constant for all marginal states. In the present case $k^2 = \pi^2/2$, implying that perturbations having a horizontal dimension roughly half that of the plate spacing will exhibit the fastest growth. It is, however, difficult to reconcile this result with the observations. The reported salt finger structure, to be discussed in the following paragraph, suggests a much larger wave number than the critical one. Stern[50] and Baines and Gill[2] both note that linear theory predicts a rapid increase in the wave number giving the fastest growth in the unstable region. Shown in Fig. 2 are the wave numbers giving the fastest growth rate and the associated exponential growth factor for states (R, R_s) which are supercritical. The particular region depicted is that in which salt fingers form. In this region the wave number associated with points in the [R, R_s] plane does increase rapidly with distance measured from the neutral stability curve.

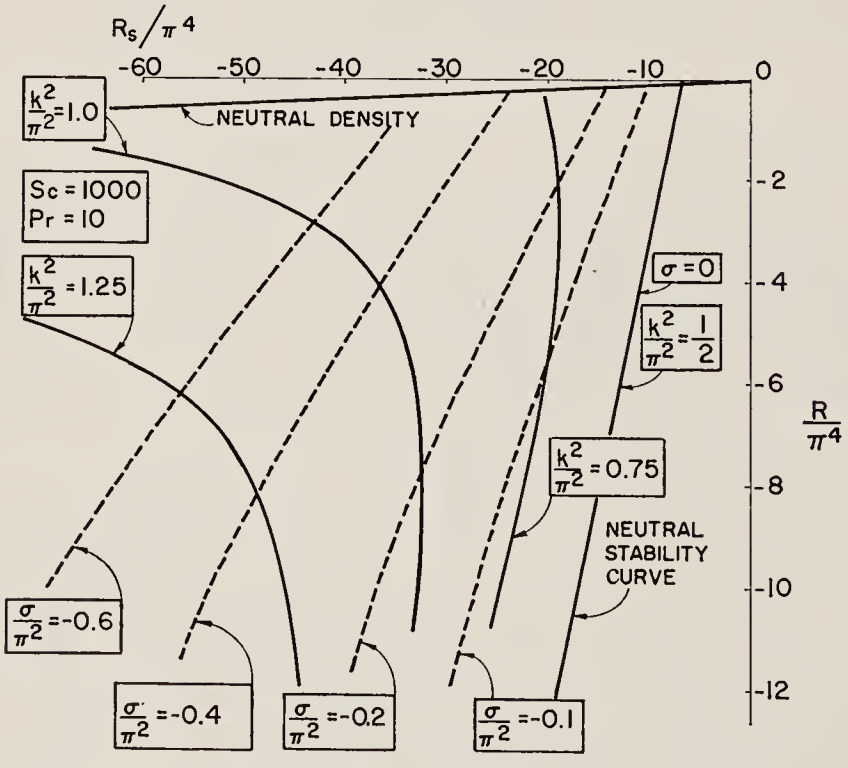


Fig. 2. The wave numbers giving the fastest growth in supercritical region and the associated exponential growth factor; calculations are based on results of linear theory.

We shall not pursue this issue further because of the difficulty of explaining nonlinear phenomena (salt fingers) using linear theory. The sharp increase of the wave number depicted in Fig. 2, however, does seem to offer a plausible explanation.

C. General Boundary Conditions

Nield[30] has considered the linear problem for a general set of boundary conditions using an approximate analysis. Numerical solutions were presented for selected cases. Although the essential facts are not different from those obtained with free boundaries, in general the neutral stability curve for stationary marginal states is no longer a straight line in the $[R, R_s]$ plane nor is the critical wave number a constant.

An interesting feature arises when boundary conditions appropriate for rigid, impermeable, and insulated walls are imposed. The critical wave number vanishes. Thus a single cell will be the favored convection pattern. Nield[30] has shown that the convection velocities associated with the single cell will be quite small, but nonzero. As has been noted by both Nield[30] and Velarde and Schechter,[63] the single cell permits convective motions with a minimum of viscous dissipation. The Rayleigh numbers required to produce such convective motions are generally quite small compared to smaller-scale flow patterns. Compare, for example, case 3 of Table I with the first two cases. As the critical wave number increases so does the critical Rayleigh number. Larger driving forces are therefore available to surmount the increased viscous dissipation.

D. Experimental Observations

The analyses presented here are all based on an assumed initial linear variation in the concentration of each diffusing component and in the temperature. In practice this initial configuration is difficult to achieve, and consequently, the prediction of linear theory has not yet been experimentally verified. There are, however, a number of reported experiments that give considerable credence to the predictions and in no way contradict them.

A number of issues are important and require experimental measurement. Of great interest is the question of existence. Do the predicted instabilities really exist or have the theorists omitted some important term leading to erroneous conclusions? If the instability does exist, do the criteria developed here correctly predict the onset of convection? What is the structure of the instability and how much energy or mass is transported by the convective motion? These questions can generally be answered only in terms of nonlinear theory (see Section VI) or by direct experimentation.

We can best understand the observations by considering those experiments related to salt fingers separately from those in which an adverse

density gradient exists. Let us consider first the salt fingers. Experiments have been conducted by Turner,[59] Stern and Turner,[52] Shirtcliffe and Turner,[47] and Lambert and Demenkow.[20] The geometry or planform of the salt fingers have been most extensively studied by Shirtcliffe and Turner. An aqueous solution of sugar was poured carefully onto a layer of more dense salt solution and the resulting convective instability was observed using a schlieren system. Three main regions were observed. Top and bottom convecting layers (reasonably well mixed) were connected to one another by fingers. The convection was thought to be mainly associated with plumes or sheets of material ejected from the finger region. These wander laterally in a haphazard manner. Only the finger region displayed a regular structure. At the onset of the experiment the interface was initially very turbulent and no regular pattern was visible. As a result of the transport of sugar and salt across the interface, the differences in concentration decreased and the motions became slower. The fingers elongated and a dominant scale appeared. The cells that formed showed a strong tendency to meet at right angles and the preferred planform was a square. During the course of a run, the fingers tended to grow in size, increasing in both length and width. Indeed they tended to grow until attaining a length about equal to the height of the containing vessel. Linear stability analysis predicts that k_c should be constant for similar boundary conditions ($k_c = 2.2$ for fixed concentrations and temperatures) which in turn implies that the width should increase linearly with increasing lengths. The observed critical wave number is far smaller than the value $k_c = 2.2$ and the width increased much more slowly than the length ($\approx d^{1/3}$). As has been noted, the explanation of this observation may simply be that in the supercritical region the value of k giving the most rapid growth is quite different from the critical value (see Fig. 2). It is, however, rather difficult and very speculative to apply these results to the transient experiments of Shirtcliffe and Turner.

The experiments performed by Turner[58] leave no doubt as to the existence of salt fingers and, moreover, suggest that the lengths of these fingers may in some way be limited. This may very well bear a relationship to the oceanographic steps underneath the Mediterranean outflow which have been recorded by Tait and Howe.[55] Well mixed layers separated by interfaces which at times were too sharp to resolve with the measurement techniques used were observed. Using a very crude order of magnitude analysis, Stern[51] has predicted that quasilaminar salt fingers can occur in a region only if the salt flux is less than a critical value given by

$$\frac{\beta F_s}{\nu\alpha(\partial T^s/\partial Z)} \cong 1 \tag{3.16}$$

Larger salt fluxes lead to an instability which would destroy the salt fingers. This criterion has not been tested explicitly but should be viewed as tentative since Lambert and Demenkow[20] have reported that the predicted mass fluxes based on Stern's approximation differ from their observations by three orders of magnitude. It does, however, appear that the $\frac{4}{3}$-law for the salt flux which was proposed by Stern and Turner[52] does correctly predict the fluxes when the constants are appropriately modified.

If we accept the concept introduced by Stern[51] that a critical salt flux does in fact exist, then one should be able to develop in the laboratory distinct layers consisting of mixing zones separated by salt fingers. Stern and Turner[52] have observed this layering phenomenon. Using a technique proposed by Oster,[31] they were able to obtain a linear gradient of the stabilizing component (salt). Subsequently a layer of less dense sugar solution was poured on top of the salt solution. With a sufficiently small concentration gradient of salt, two layers were observed to form with quasilaminar regions of stable salt fingers connecting the layers. Thus a partial verification of (3.16) is obtained. Its complete verification must await more precise experiments in which the gradients and attendant fluxes are measured simultaneously.

Salt fingers occur when the density gradient is not adverse. For adverse density gradients produced by heating from below but with a stabilizing solute configuration, an overstable response is predicted. Shirtcliffe[44,45] and Turner[59] have observed the overstability. Furthermore it now seems clear that the very interesting observations reported by Miller et al.[28,29] are in fact examples of the overstability under discussion. They observed an oscillatory behavior during a study of three-component diffusion in which the diffusional fluxes were oriented parallel to gravity.

The most extensive attempt to verify the predicted limits of marginal stability is the work of Shirtcliffe.[45] He found that in all observed cases convection did indeed begin at the marginal stability with an oscillatory motion as predicted by linear theory. The boundary of stability in the (R, R_s) plane was observed to depart somewhat from that predicted by the simplest theory (Section III.A); the boundary is a curve, concave to the origin and the critical Rayleigh number is high. The curvature of the neutral stability curve is to be expected if the boundary conditions are no longer "similar" (Section IIIC). Since the experiments conducted by Shirtcliffe do not model the mathematical development (the imposed temperatures varied with time), it is not surprising to note a variation between the observed and predicted critical Rayleigh number.

Shirtcliffe was able to provide convincing experimental evidence that the period of the oscillations is in reasonable agreement with the predictions.

Turner[59] studied the convective motions produced by heating from below a solution initially stabilized by a linear salinity gradient. He observed the formation of a convecting layer whose depth, temperature, and concentration all increased as $(t)^{1/2}$. After a certain critical time elapsed, the upper boundary of the deepening layer seemed to stagnate, and a second convecting layer would then spontaneously form on top of the stagnant one. Shirtcliffe[46] has confirmed this behavior. Turner[59] developed a stability criterion which predicts the critical time as a function of the relevant parameters. To date there have not been reported careful experiments which correspond to the initial configurations assumed in the theoretical calculations.

IV. CONVECTIVE MOTION AND THERMAL DIFFUSION: LINEAR THEORY

A. The Temperature Coupled Mass Balance

Agar and Turner[1] observed that Soret coefficients for certain electrolytes were consistently larger when measured in a transient experiment than in a steady-state one. Even though these experiments were all conducted in cells which were heated from above and therefore expected to be stable, they proposed an index of stability on more or less empirical grounds. Whether they suspected the existence of convection currents or merely a greater tendency for convection which would aggravate experimental problems such as the alignment or minimizing heat losses, their index, as we shall see, has considerable merit. Having examined "salt fingers," we may already suspect possible instabilities even for systems heated from above.

The mass flux of the more dense component in a binary mixture is given by (see Ref. 11, p. 276)

$$\mathbf{j}_1 = -\rho S_1 S_2 D' \nabla T - \rho D \nabla S_1 \tag{4.1}$$

where D' is the thermal diffusion coefficient. The ratio D'/D is called the Soret coefficient. To measure the Soret coefficient it is usual to subject binary fluid contained between horizontal plates to a temperature gradient with the heat applied from above so as to prevent convection. At the steady state the separation owing to thermal diffusion is balanced by diffusion and the flux of mass vanishes. Thus in dimensionless terms

$$\frac{\partial S_1}{\partial Z} = -\frac{D' S_1 S_2}{D} \frac{\partial T}{\partial Z} \tag{4.2}$$

Since D'/D is generally quite small, ranging from 10^{-3} to $10^{-5}°C^{-1}$ for both liquids and gases, the separations attending thermal diffusion are

TABLE II
Migration Direction of the More Dense Component

	$D' < 0$	$D' > 0$
$\Delta T > 0$	*Quadrant IV:* More dense component migrates to warm lower boundary	*Quadrant I:* More dense component migrates to cold upper boundary
$\Delta T < 0$	*Quadrant III:* More dense component migrates to warm upper boundary	*Quadrant II:* More dense component migrates to cold lower boundary

correspondingly small. In the calculations to follow we take S_1S_2 to be constant and equal to an average value $S_1{}^*S_2{}^*$. This assumption is questionable only in very dilute solutions. For a linear temperature gradient

$$\Delta S_1 = \frac{-D'S_1{}^*S_2{}^*\,\Delta T}{D} = -\frac{P\,\Delta T}{D} \tag{4.3}$$

Recalling that $\Delta[\ ] = [\ ]_{\text{lower}} - [\ ]_{\text{upper}}$, the results shown in Table II are easily deduced.

Substituting the mass flux expression, (4.1), into (2.13), the following linearized mass balance results:

$$\sigma\Gamma_1 + W\,\Delta S_1 + \frac{1}{Sc}\left[\frac{d^2\Gamma_1}{dZ^2} - k^2\Gamma_1\right] + \frac{S_1{}^*S_2{}^*D'\,|\Delta T|}{\nu}\left[\frac{d^2\Theta}{dZ^2} - k^2\Theta\right] = 0 \tag{4.4}$$

This expression differs from the mass balance of the salt finger problem since here there is a direct coupling between concentration and temperature fluctuations. Equation (2.15) applies if the Dufour effect is neglected. This assumption is justifiable for liquid systems (see Ref. 11).

It is convenient to define a new variable

$$\tilde{\Gamma} = \Gamma_1 + P\,|\Delta T|\,\Theta \tag{4.5}$$

which has the property $d\tilde{\Gamma}/dz = 0$ at an impermeable boundary. In terms of this new variable, the mass balance is

$$\sigma\tilde{\Gamma} - \sigma\frac{P\,|\Delta T|}{D}\,\Theta - \frac{P\,\Delta T}{D}\,W + \frac{1}{Sc}\left[\frac{d^2}{dZ^2} - k^2\right]\tilde{\Gamma} = 0 \tag{4.6}$$

The momentum and energy balances appropriate to the salt finger problem apply here, and if the fluid is contained between impermeable conducting boundaries (the most realistic approximation), then

$$W = \frac{dW}{dZ} = \Theta = \frac{d\tilde{\Gamma}}{dZ} = 0 \qquad \text{at} \qquad Z = 0, 1$$

The solution to this problem is complex, and considerable insight can be gained by studying the physically unrealistic but mathematically simpler problem of free boundaries defined by the boundary conditions

$$W = \frac{d^2W}{dZ^2} = \Theta = \tilde{\Gamma} = 0 \qquad \text{at} \qquad Z = 0, 1$$

B. Free Boundaries

For this case the eigenfunctions again have the form $[W, \Theta, \tilde{\Gamma}] = [A, B, C] \sin n\pi z$ and the solution is nontrivial if

$$q^2 - q^2\left(1 + \frac{1}{Sc} + \frac{1}{Pr}\right) + q\left(\frac{1}{Pr} + \frac{1}{Sc} + \frac{1}{ScPr} - R'\right) + \frac{R'}{Sc} - \frac{R_s'}{Pr} - \frac{1}{PrSc} = 0 \quad (4.7)$$

with

$$R' = \frac{k^2}{a^6}\left[\frac{R}{Pr} - \frac{R_s}{Sc}\right]$$

$$R_s' = \frac{k^2}{a^6}\frac{R_s}{Sc}$$

Overstable marginal states are found when $q = i\omega$ (ω real). Equation (4.7) is satisfied for this case if

$$\omega^2 = \frac{1}{Sc^2} + \frac{R_s'}{Pr}$$

The most sensitive mode is the fundamental with $n = 1$ and $k = \pi/\sqrt{2}$. We then find ($\sigma = \sigma_R + i\sigma_I$)

$$\sigma_I^2 = -\frac{9\pi^4}{4Sc^2} + \frac{R_s}{3Sc(1 + Pr)} \quad (4.8)$$

and

$$R_c^{ov} = \frac{27\pi^4}{4}\left(1 + \frac{Pr}{Sc}\right)\left(1 + \frac{1}{Sc}\right) - \frac{PrR_s}{Sc(1 + Pr)} + \frac{PrR_s}{Sc} \quad (4.9)$$

Since both R and R_s increase with increasing temperature, the results can be made more transparent by defining a third dimensionless group

$$S = -\frac{PrR_s}{ScR} = \frac{\beta P}{\alpha D} \tag{4.10}$$

This quantity relates the density change owing to thermal diffusion to that resulting from temperature changes of constant composition. Thus

$$\frac{\partial \rho}{\partial Z} = \rho_m \alpha \Delta T(1 + S) \tag{4.11}$$

Equation (4.9) can be rearranged to yield

$$R_c^{ov} = \frac{27\pi^4}{4}\left[\frac{(1 + Pr)(Sc + Pr)(1 + Sc)}{Sc^2(PrS + Pr + 1)}\right] \tag{4.12}$$

a result first given by Hurle and Jakeman.[14] There are no physical restrictions on either the sign or the magnitude of S. For positive R, the overstable region is restricted to values of S satisfying the inequality

$$\frac{1 + Pr}{Pr} > -S > -S^* = \frac{Pr(1 + Pr)}{(1 + Sc)(Sc + Pr) + Pr^2} \tag{4.13}$$

When these conditions are met, oscillations should be observed since $R_c^{ov} < R_c^{st}$ where the Rayleigh number for a stationary marginal state is given by the condition $q = 0$ and found to be

$$R_c^{st} = \frac{27\pi^4}{4}\frac{Pr}{Pr + S(Sc + Pr)} \tag{4.14}$$

The curves defining neutral stability are shown in Fig. 3. The quadrants correspond to those of Table II so that measurements of the Soret coefficients are normally conducted in Quadrants II and III. In Quadrant III a stationary instability is possible even though the density gradient as expressed by (4.11) is negative.

C. Impermeable Conducting Boundaries

This model can be very nearly realized in the laboratory and is therefore of the greatest interest. Unfortunately a complete solution for both stationary and oscillatory marginal states is not yet available. Schechter, Prigogine, and Hamm[40] have obtained an exact solution for stationary marginal states, but offered only approximate treatment of overstability. Hurle and Jakeman[15] and Legros, Platten, and Poty[26] have also presented approximations defining the neutral stability curves. Although all the results are in reasonable agreement, an exact definition of neutral stability for

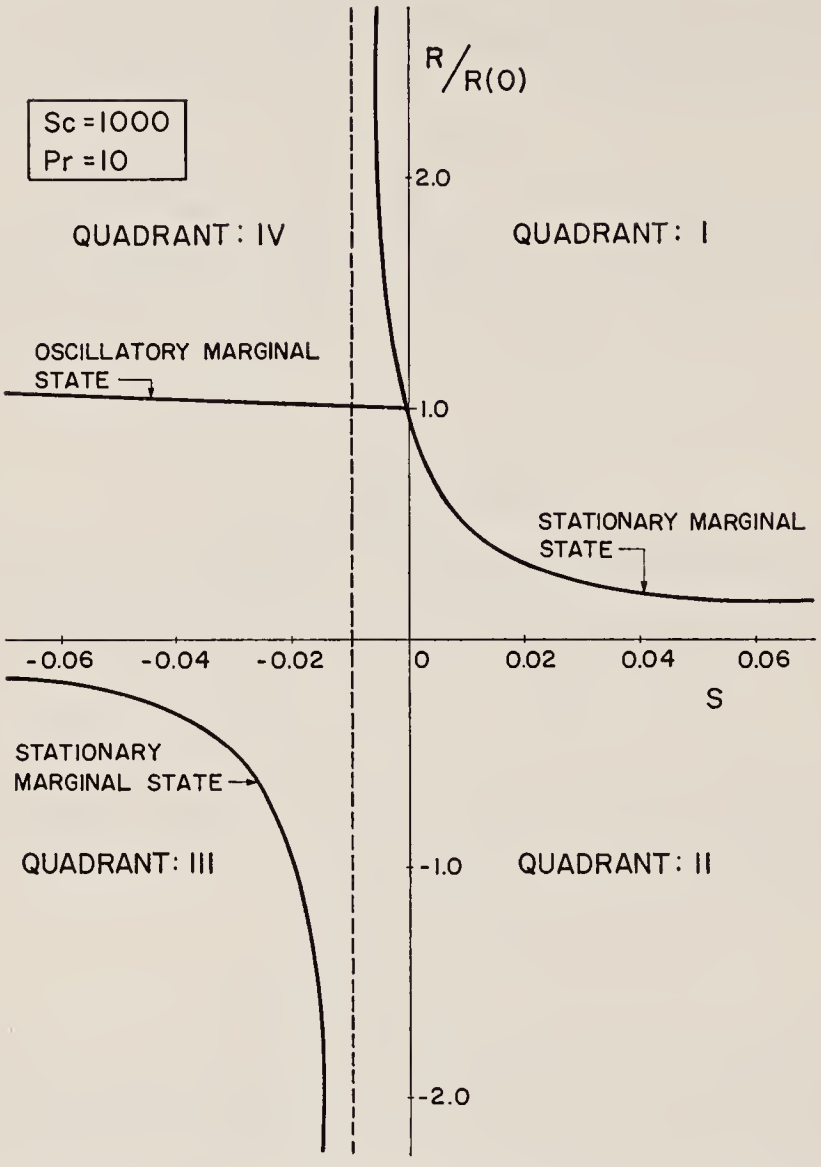


Fig. 3. Neutral stability curves for the thermal diffusion problem. The quadrant numbers refer to Table II.

overstable states would help to resolve certain difficulties which will be discussed.

There are interesting differences between the results for this model and those found for the free boundary case. Shown in Fig. 4, for example, is the variation of the critical wave number, k_c, as a function of H where

$$H = \frac{ScS}{Pr(1 + S)}$$

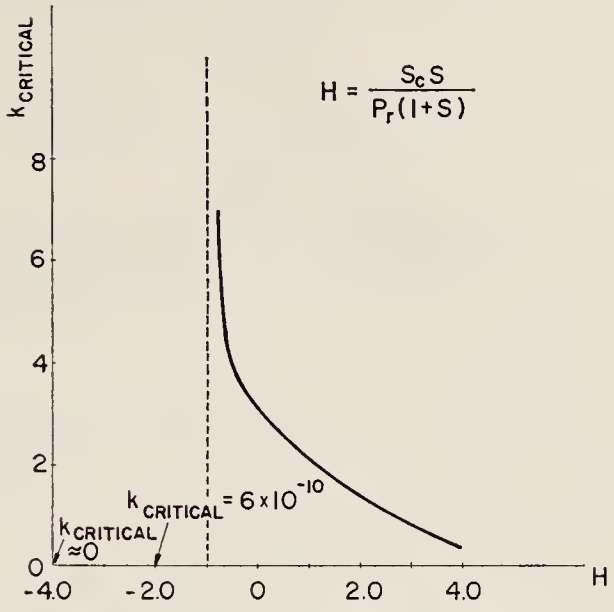


Fig. 4. The critical wave number for impermeable, conducting rigid boundaries. $H = ScS/Pr(1 + s)$.

As H increases, k_c rapidly approaches zero in sharp contrast with the free boundary case in which k_c is a constant. For $H \leftarrow 1.0$, k_c is also very nearly zero. This rapid increase in the size of a cell provides a mechanism for sustaining convection with the minimum of viscous dissipation. The same driving forces are operative at both small and large values of H. This is verified in the following section.

For values of H between the limits

$$-1 < H < H^*$$

the critical Rayleigh number is, as was the case for free boundaries, greater than the overstable one. Thus in this region an oscillatory state will be observed. Platten and Chavepeyer[33] have suggested that H^* can be found using an expression valid for all boundary conditions, namely,

$$S^* = \frac{-CPr(1 + Pr)}{(1 + Sc)(Sc + Pr) + Pr^2} \tag{4.15}$$

where C is a constant that depends on the boundary conditions. Of course, $C = 1$ for free boundaries, but since other exact solutions are not available, we can only infer the validity of this expression from approximate numerical results.

The neutral stability curve for the special case $Sc/Pr = 100$ is shown in Fig. 5. The similarities between this curve and that given in Fig. 3 for the case of free boundaries are obvious. The overstable marginal state is

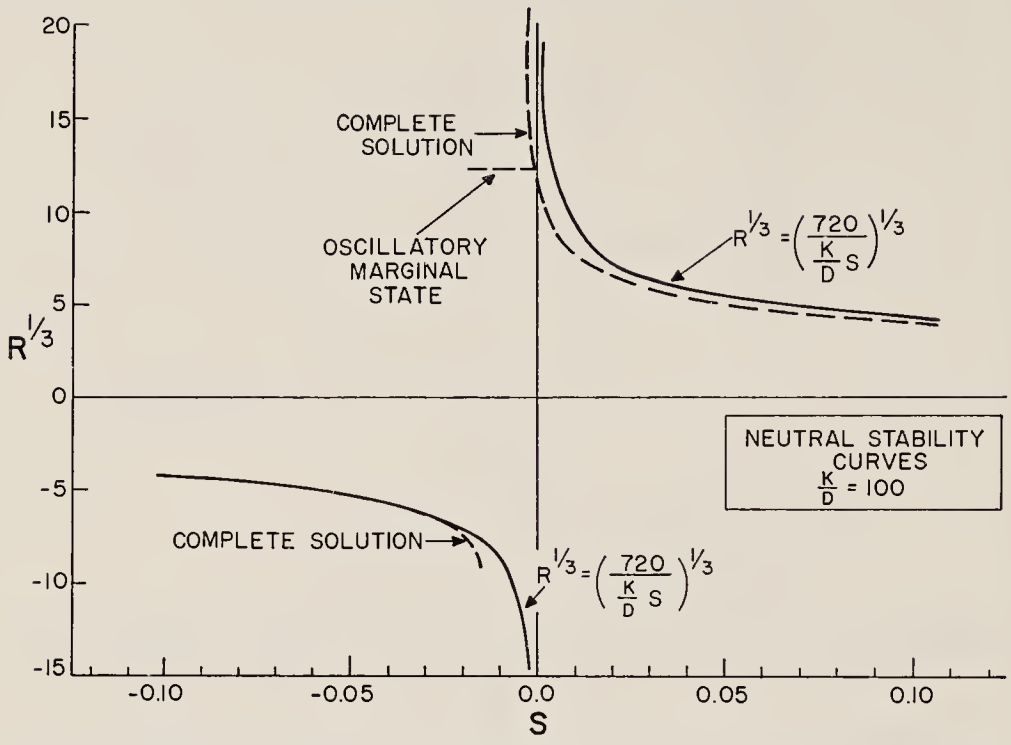


Fig. 5. Neutral stability curves for impermeable, conducting, rigid boundaries. The oscillatory marginal state is only approximately known. The dashed line represents the complete solution of Schechter et al.[40]

taken from the work of Hurle and Jakeman;[15] however, their approximation is suspect since it erroneously predicts that for certain small values of S stationary convection is possible when the system is heated from above or from below.

Finally, Platten and Chavepeyer[34] suggest that the frequency of the overstable oscillations is given by the equation

$$\sigma_I^2 = -C_1 \frac{S}{Pr(1 + Pr)} - C_2\left(\frac{1}{Sc}\right)^2$$

where C_1 and C_2 depend on the boundary conditions. For the present case $C_2 = 93.5492$ and $C_1 = 0.167629$ have been suggested instead of $C_2 = 219.17046$ and $C_1 = 0.333333$ for free boundaries. These coefficients have been obtained by curve-fitting numerical results. We note that for a typical case $S = -0.5$, $R_c^{ov} \cong 3300$, $S_c \cong 1000$, and $Pr \cong 10$, so that $\sigma_c^2 \cong 2.51$ and the period is

$$\Phi = \frac{2\pi}{(\nu/d^2)\sigma_c}$$

Thus for the reasonable depth of 0.4 cm,

$$\Phi \cong 63 \text{ sec}$$

which is readily detectable.

D. An Approximation

Velarde and Schechter[63] have expanded the velocity, temperature, and concentration profiles in powers of Pr/Sc. The first approximation is

$$\Theta = 0 \tag{4.16}$$

$$\left[\sigma + \frac{d^2}{dZ^2} - k^2\right]\tilde{\Gamma} + \Delta S_1 W = 0 \tag{4.17}$$

$$\frac{R_s}{\Delta S_1}\tilde{\Gamma} + \left[\frac{d^2}{dZ^2} - k^2\right]^2 W = 0 \tag{4.18}$$

Imposing the boundary conditions applicable to an impermeable boundary, σ can be shown to be real, and therefore in contrast with the general situation, overstable states are not found in the first approximation.

Because of the similarity between this mathematical problem and Case 10 of Table I, we can immediately write

$$-R_s = R_c^{st} S \frac{Sc}{Pr} = 720 \tag{4.19}$$

and

$$k_c = 0 \tag{4.19}$$

This approximation is compared with the complete nonoscillatory solution and the result is shown in Fig. 5. As can be seen, for sufficiently large S, the approximation is quite accurate, and the temperature fluctuations do not, therefore, contribute to the instability [see (4.16)]. Furthermore the onset of convective motion cannot be detected by observing the heat flux which is not changed by the initiation of large-scale ($k_c = 0$), very slow motions.

E. Experimental Results

We have seen that a very unexpected instability develops when, on heating a binary fluid contained between horizontal parallel plates from above, the more dense component migrates to the warm upper wall. This convective instability can be expected to interfere with the measurement of the Soret coefficient especially if the experimentalist is unaware of the danger. Many of the experimental results relating to this instability have been obtained inadvertently.

*Quadrant I** ($R > 0, S > 0$)

In this quadrant the thermal diffusion reinforces the adverse density gradient established by the imposed temperature gradient. The critical Rayleigh number is thereby significantly reduced. Generally a Schmidt-Milverton[41] plot of heat flux as a function of temperature difference does not reveal the onset of convection. If, however, a plot of mass separation as a function of ΔT was constructed, the instability should appear as a change in slope. Such an experiment has not yet been reported. Legros et al.[22] sought to find the critical Rayleigh number for benzene-carbon tetrachloride and chlorobenzene-carbon tetrachloride mixtures without success, because they relied on Schmidt-Milverton plots to reveal the unstable point. Legros et al.[23] did sample the contents of a diffusion cell and found no concentration separation for a $C_2H_2Br_4$-$C_2H_2Cl_4$ mixture at temperature differences which were 0.2% of the usual Bénard temperature gradient. Using the data presented by Legros et al.[23] and estimating some of the physical properties, $S \cong 0.8$ and $Sc/Pr \cong 100$. Thus $R_c^{st} \cong 9$ or about 0.5% of the 1708 for the pure component problem. If the uncertainties in the physical properties are considered, this computed value is in reasonable agreement with the observation.

Quadrant II ($R < 0, S > 0$)

Most measurements of the Soret coefficient are taken in this quadrant. In a well-managed carefully performed experiment there should be no convection.

Quadrant III ($R < 0, S < 0$)

Belton and Tyrrell[6] measured the mass separation within a toluene-ethanol layer subjected to a negative temperature gradient (heated from above) and did not observe any anomalies. Since these experiments were almost certainly conducted in the stable region, they represent a partial confirmation of the theory.

If small amounts of methanol are mixed with benzene, the resulting solution has a negative S. This system has been studied by Thomaes,[56] Tichacek et al.,[57] Whitaker and Pigford,[67] Tyrrell et al.,[60,61] and Story and Turner.[54] As noted by Velarde and Schechter,[62] the magnitudes of the reported Soret coefficients differ widely among these investigators and are definitely arranged so that the ones with the largest values of $-RS$ report the smallest Soret coefficients. Of course when these data were obtained, the concept of an instability in the face of a stable density gradient seemed absurd.

* See both Table II and Fig. 5.

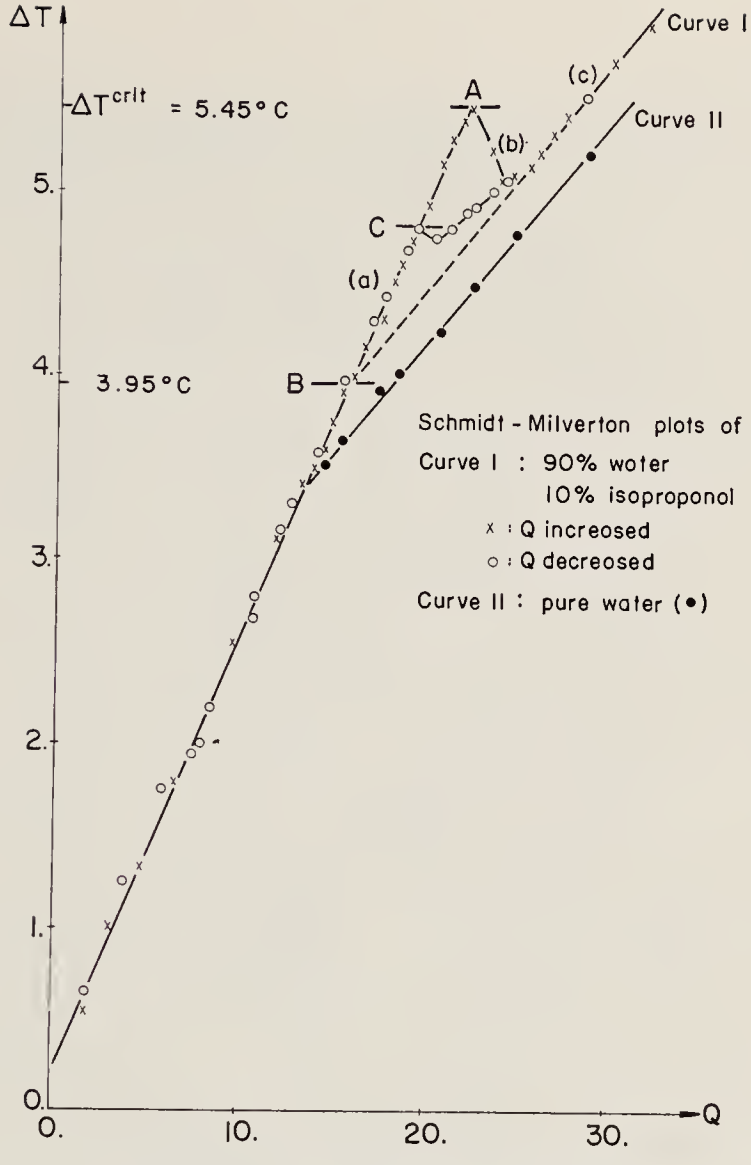


Fig. 6. A plot of imposed temperature difference as a function of heat flux. The results are those of Platten and Chavepeyer.[34]

Quadrant IV ($R > 0$, $S < 0$)

In this quadrant thermal diffusion can be expected to contribute a stabilizing influence. This stabilization has been observed by Caldwell,[7] Legros et al.,[25] and Platten and Chavepeyer.[34] The Schmidt-Silverton plot for a water-isopropanol mixture (10 wt % propanol] shown in Fig. 6 is typical. When the heat flux is small, corresponding to a small ΔT, then the heat is transferred through the fluid layer by conduction and the slope is related to the thermal conductivity. Increasing the heat flux gives rise to an increase in the temperature gradient until point A is reached. At

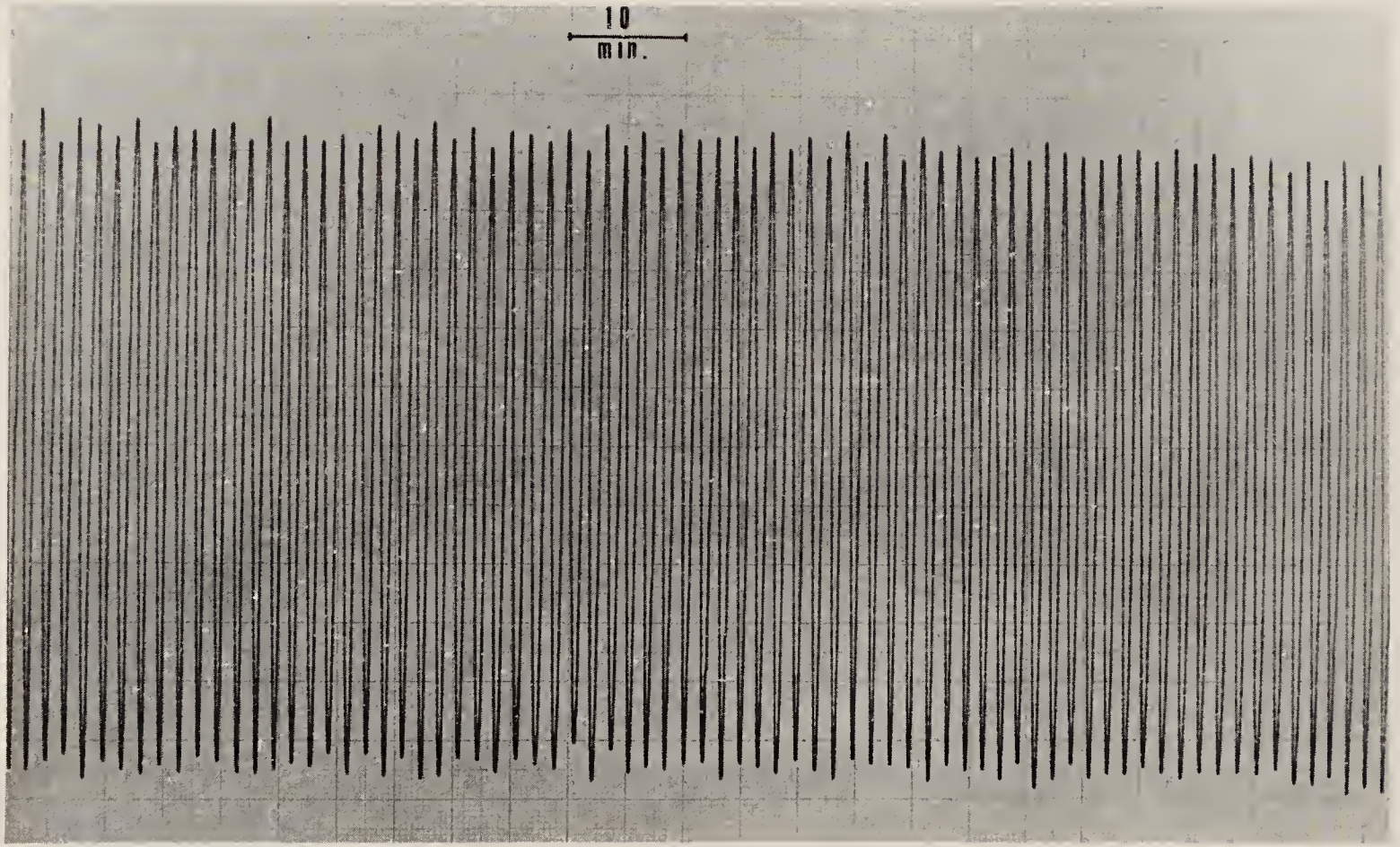


Fig. 7. Sustained temperature oscillations observed in water-isopropanol (10 wt %) layer heated from below; taken from Platten and Chavepeyer.[34]

this point oscillatory convection currents are formed and a negative-slope region is seen. The oscillations apparently subside at sufficiently large heating rates, and a curve reminiscent of that observed using pure fluids is obtained (see curve II in Fig. 6 which is for water). In this region the curve is again linear, having a different slope because of the contribution of convection to the heat flux. It is fascinating to note that if the curve is retracted by slowly decreasing the heat flux, there is a hysteresis giving a different curve until the conduction branch is intersected.

If point B is found by extrapolation, then the distance from A to B can be considered a measure of the stabilizing influence of thermal diffusion. Caldwell[8] has successfully utilized this concept to obtain the Soret coefficients of NaCl in water at mean temperatures such that S is negative.

Both Hurle and Jakeman[15] and Platten and Chavepeyer[34] have observed oscillations in the negative slope region of the Schmidt-Milverton plot. Hurle and Jakeman were not able to stabilize the oscillations, perhaps because their experiments were conducted utilizing a continuously increasing power input. Platten and Chavepeyer have, on the other hand, observed sustained oscillations. Fig. 7 shows the observed temperature oscillations for the system water-isopropanol (90:10 wt %). These appear to be quite regular and not damped. Thus the oscillations in the negative-slope region may not grow spontaneously until a finite-amplitude stationary mode is excited, as has been suggested by Shirtcliffe[45,46] and Hurle and Jakeman. The issue is not yet resolved, however, because Hurle and Jakeman[16] have shown that steady sustained oscillations can be produced even in pure fluids if the upper surface of the confining cell is sloped with respect to the horizontal by as little as two degrees. The hydrodynamic mechanisms operative in the negative slope region are not yet clearly defined.

V. CHEMICAL INSTABILITIES

Chemical reactions can promote hydrodynamic instability. Indeed we would expect that the convective instabilities described in Section III will be observed within a heated fluid layer of reacting fluids, provided the rate of chemical reaction is "slow." If, however, the persistence of a concentration fluctuation is greatly reduced by the chemical reaction, then these mechanisms giving rise to salt fingers or overstability are no longer operable and the question of stability is an open one. Wollkind and Frisch[68] studied a fluid in a chemical quasiequilibrium state, that is, one which is always very near equilibrium locally (see Lighthill[27] for a discussion of this model). The particular kinetic expression used by Wollkind

and Frisch is

$$\frac{1}{\rho_m} J_1 = \frac{S_1 - S_1^*}{\lambda} \tag{5.1}$$

where S_1^* is the equilibrium weight fraction of one component and λ is the relaxation time. For the quasiequilibrium case, λ is small, and therefore the rate of chemical reaction fast. Using (5.1) and (2.3) and allowing for the heat of reaction, it can be shown that for suitable boundary conditions the principle of the exchange of stabilities is valid and overstable cases do not exist. Furthermore, for free boundaries with specified temperatures and concentrations

$$R_c^{st} = \frac{27\pi^4}{4}\left(1 + \frac{3\pi^2}{2}\epsilon\beta + 0(\epsilon^2\beta^2)\right) \tag{5.2}$$

where

$$\epsilon = -\frac{\lambda v}{d^2}\frac{S^*(1 - S^*)}{(2 - S^*)}\frac{D}{k}\frac{\beta}{\alpha}$$

$$B = \frac{1}{Sc} - \frac{1}{Pr}$$

k = Boltzmann's constant

The critical Rayleigh number defined by (5.2) reduces to that of a pure fluid as $\lambda \to 0$. For a system heated from above

$$R_c^{st} = -\frac{4}{\epsilon^2 B^2} - \frac{2\pi^2}{\epsilon\beta} + 0(1) \tag{5.3}$$

for $B > 0$. Thus a stationary instability can exist in a reacting system heated from above provided $B > 0$ or, therefore, $Sc < Pr$. In this case the thermal fluctuations must persist longer than the disturbances in concentration. Since $B < 0$ for liquids, this instability can exist only in gases, and the roles of thermal and concentration fluctuations are reversed from the usual salt fountain case, just as would be anticipated in the case of fast reactions.

Bdzil and Frisch[5] studied the stability of more general reaction mechanisms, and their results show that for a certain range of reaction rates, instabilities closely resembling the salt case can arise. Such instabilities are possible only if the reactants are far from equilibrium, implying that the reaction rates are slow.

Bdzil and Frisch[3,4] have also studied the effect of surface reactions on hydrodynamic stability. This problem is of less interest than the homogeneous case since a surface reaction is formulated by appropriately

modifying the boundary conditions, but the differential equations correspond precisely to those used to describe thermohaline convection. Therefore, except for some variation in the critical wave number and critical Rayleigh numbers, we would expect these results to be the same as those described in Section III. We will not give further details here, therefore.

VI. GLOBAL STABILITY: ENERGY METHODS

Necessary conditions for stability can be defined by studying the fate of small disturbances about the state whose stability is in question. Thus if a state is unstable with respect to small perturbations, it is certainly unstable. The question of stability with respect to arbitrary perturbations still remains, however, even when the linear analysis is completed. Liapunov stability methods (see LaSalle and Lefschetz[21]), on the other hand, provide sufficient conditions for the stability of a certain system state. The Liapunov method that we describe here is quite simple in concept but difficult in execution. An "energy" is defined (Joseph,[17] Joseph,[18] Shir and Joseph,[43] Serrin[42]) having the property

$$E \geqslant 0 \tag{6.1}$$

where the equality applies if and only if the system is in the basic state. If it is possible to prove that

$$\frac{dE}{dt} < 0 \tag{6.2}$$

for all possible disturbances, then the basic state is certainly a stable one. Of course, we must recognize that the state in question can be a stable one even if dE/dt is not negative definite. If, for example, the decay of a perturbation was an oscillatory one, then inequality (6.2) would not be satisfied and yet the system may return to the state, $E = 0$.

We should note that when the limits of linear stability (necessary conditions) correspond to the sufficient conditions for global stability, then these limits are both necessary and sufficient guaranteeing the nonexistence of finite amplitude subcritical instabilities. Sani[37] has shown that subcritical instabilities can not exist when a chemically homogeneous fluid layer is heated from below. If concentration gradients exist, the situation is not so well defined.

Shir and Joseph[43] defined energy which is simply a weighted sum of $\langle \mathbf{v}^2 \rangle$, $\langle (T - T^*)^2 \rangle$, and $\langle (S_1 - S_1{}^*)^2 \rangle$, where T^* and $S_1{}^*$ are the temperature and concentration distributions whose state is to be tested for stability. The brackets denote an integration over the volume of the system, and the weighting was selected to obtain the widest possible limits of stability for the particular Liapunov function. A universal criterion for stability

was developed and can be written as follows:

$$\sqrt{2368} > \sqrt{|R|} + \sqrt{|R_s|} \tag{6.3}$$

All states (R, R_s) satisfying inequality (6.3) are stable. States violating this inequality may or may not be stable. This result is independent of the Sc and Pr numbers as well as the signs of R and R_s or the boundary conditions.

A second important result given by Joseph[18] is that

$$\frac{Pr}{\sqrt{Sc^2 - Pr^2}}\left(\frac{Sc}{Pr}\sqrt{|R|} - \sqrt{|R_s|}\right) < \sqrt{R^*} \tag{6.4}$$

where R^* is the critical Rayleigh number for a homogeneous fluid under the same boundary conditions. This inequality is restricted to cases for which

$$\frac{|R|_s}{|R|} > 1 \qquad \text{and} \qquad \frac{Sc}{Pr} > 1$$

Inequality (6.4) is, in a sense, an extension of inequality (6.3) in that the role of both the Schmidt and Prandtl numbers is clearly shown. The boundary conditions also are included in this result since R^* depends on them.

Subcritical finite amplitude instabilities are ruled out for any state satisfying inequality (6.4). A comparison with this condition and those developed in Section III will establish the conservative nature of predictions based on inequality (6.4). Nevertheless, this equality is powerful since regions for which finite amplitude instabilities are impossible are clearly delineated.

VII. SUPERCRITICAL STATES

A. A Classification of Problems

Convective instabilities, once triggered, give rise to enhanced transport of both mass and heat, since the system evolves from a quiescent state to one of motion. As has been noted, the new supercritical state may be stationary or oscillatory in time. Indeed the first observed motions may be oscillatory ones which subsequently lead to stationary states. Such a series of events would be observed, if the oscillatory motions are themselves unstable and grow with time. The response to questions such as "what is the planform of the supercritical state?" or "what are the new fluxes of mass and energy?" can be found only by including the nonlinear terms appearing in (2.1), (2.2), and (2.3) in the analysis. To address these questions, it is more convenient to study two cases separately. The first is the thermohaline convection with a stabilizing solute gradient. Overstability is

predicted for the onset of convection as is shown by Fig. 1, which is valid for free boundaries. This problem has been studied by Sani[38] and Veronis.[64,65] The second case is a destabilizing solute gradient. As we shall see, if this gradient is sufficiently large, the instability which occurs at small positive or negative Rayleigh numbers has a structure that is almost independent of the direction of the temperature gradient. For smaller solute gradients, the convection triggered by heating from below will closely resemble that of the classical Bénard problem and will not be considered here.

B. Stabilizing Solute Gradients

Both Sani[38] and Veronis[64,65] consider the case of free boundaries and both assume the preferred planform to be two-dimensional rolls. Although two-dimensional rolls have not been observed in the reported experiments, neither has any other distinct planform been seen to form. In viewing the photographs taken by Shirtcliffe[45,46] some structure can be detected, but these are not sufficiently distinct to indicate a preferred one.

Once a planform is assumed, then the supercritical states can be represented in terms of Fourier series with time-dependent coefficients. The time dependence of these coefficients is determined from the differential equations using what amounts to a Galerkin procedure (see Schechter[39]). This procedure leads to an infinite set of coupled differential equations, one for each unknown Fourier coefficient. To obtain numerical results, the Fourier series must be truncated. Veronis[64] and Sani[38] have retained one or two terms in their analysis, but Veronis[65] was able to retain more terms and obtained convergence for a limited range of ratios of the Prandtl to Schmidt numbers.

Veronis[65] reports that his calculations indicate an oscillatory instability for $R_s = 1000$ and $Pr = 1.0$ in the range $1797 \leqslant R < 2000$ and a well-developed, finite-amplitude steady motion at $R = 2000$. Since the point of instability to overstable modes is $R_c^{ov} = 1797$, the fluid first becomes unstable to overstable modes and these persist, increasing in amplitude as R increases. When R reaches a value between 1900 and 2000, a steady convective pattern develops. This behavior was suggested by the restricted region shown in Fig. 1 in which linear theory suggests that oscillatory motions are permitted. This passage from an oscillatory to a stationary state has been partially verified by experiments. The relevant results are described in Section III.D. The weakness of the experimental results stems from the difficulty in maintaining appropriate solute concentrations at a boundary. Indeed all reported experiments are inherently time dependent in that there has been no provision made to restore the solute gradient which is destroyed by convective mixing.

Veronis[65] and Sani[38] have both verified the greater efficiency of stationary motion in convecting both energy and matter so that the transition from oscillatory to stationary motions should be observable in terms of an increased heat flux. This greater rate of heat transfer accounts for the negative slope region of the Schmidt-Milverton plot shown in Fig. 6. Upon reaching point A the onset of instability is characterized by an oscillatory convective motion which upon further increase in the heat flux evolves to a state of stationary motion.

C. Destabilizing Solute Gradients

If the salt gradients are stabilizing, the flow behavior is complex in the supercritical region with the motions first being oscillatory and then becoming stationary upon modest increases in the Rayleigh number. With a destabilizing solute gradient, the situation is quite different especially for larger solute gradients. Consider, for example, the thermal diffusion case. With a destabilizing solute gradient ($R > 0$, $S > 0$ and $R < 0$, $S < 0$), only stationary instabilities exist. Furthermore, the critical wave number becomes small whether the heating is from above or below (see Fig. 4), and the instability in the region of adverse buoyancy ($R > 0$, $S > 0$) closely resembles that in the region of nonadverse buoyancy ($R < 0$, $S < 0$). Indeed the approximate results obtained by Velarde and Schechter[63] are in excellent agreement with more precise calculations of the neutral stability, as is depicted in Fig. 5, as long as S is sufficiently large. Furthermore, the approximation correctly predicts (see Section IV.D) that the same stability criterion, (4.19), applies to both regions. This correspondence between the approximate solution and the complete linear stability analysis implies that the expansion in powers of Pr/Sc is valid. Furthermore, the first approximation contains most of the essential features of the linear problem. It is perhaps reasonable to expect that the expansion temperature, solute concentrations, pressure, and velocity appearing in (2.1), (2.2), and (2.3) in powers of Pr/Sc will yield approximate nonlinear equations which are simplified and therefore tractable.

Velarde and Schechter[63] have developed the expansions suggested above and have for the limiting case of $k \to 0$ obtained mass flux expressions which predict the observations of Agar and Turner.[1] These results are not given here because it is now clear that k_c will be small only for states very close to the marginal state and that in the supercritical region a structured planform resembling fingers will form as has been described. The results of the Velarde and Schechter analysis are therefore not likely to be generally applicable. However, the heat flux does vanish for all values of k in the first approximation.

Acknowledgment

The authors wish to thank Professor I. Prigogine for the ample guidance and thoughtful insight he has provided during the course of our studies. One of the authors (Robert S. Schechter) has received financial support from the National Science Foundation (NSF Grant ≠ GK-11562) and wishes to express his appreciation of that help; acknowledgment should also be given to NATO (Research Grants Programme n°644) making a visit to Brussels possible.

References

1. J. N. Agar and J. C. R. Turner, "Thermal Diffusion in Solutions of Electrolytes," *Proc. Roy. Soc. London Ser. A*, **255**, 307 (1960).
2. P. G. Baines and A. E. Gill, "On Thermohaline Convection with Linear Gradients," *J. Fluid Mech.*, **37**, 289 (1969).
3. J. Bdzil and H. L. Frisch, "Chemical Instabilities II: Chemical Surface Reactions and Hydrodynamic Instability," *Phys. Fluids*, in press (1973).
4. J. Bdzil and H. L. Frisch, "Chemical Instabilities IV: Non-Isothermal Chemical Surface Reactions and Hydrodynamic Stability," *Physics* of *Fluids*, in press (1973).
5. J. Bdzil and H. L. Frisch, "Chemical Instabilities VII: A Note on the Hydrodynamic Stability of the Dissociating Fluid $A_2 \rightleftarrows 2A$," *Physics* of *Fluids*, in press (1973).
6. P. S. Belton and H. J. V. Tyrrell, "Thermal Diffusion in a Horizontal Liquid Film Heated from Below," *Chem. Phys. Lett.*, **4** (1970).
7. D. R. Caldwell, "Non-linear Effects in a Rayleigh-Benard Experiment," *J. Fluid Mech.*, **42**, 161 (1970).
8. D. R. Caldwell, "The Measurement of Negative Thermal Diffusion Coefficients by Observing the Onset of Thermohaline Convection," submitted for publication, *J. Phys. Chem.* (1973).
9. S. Chandrasekhar, *Hydrodynamics and Hydrodynamic Stability*, Chap. 2, Clarendon, Oxford, 1961.
10. M. Gregg, "The Microstructure of the Ocean," *Sci. Am.*, **228**, 64 (1973).
11. S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics*, North-Holland, Amsterdam, 1962, pp. 273–284.
12. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability, and Fluctuations*, Chapts. V and VI, Wiley-Interscience, New York, 1971.
13. I. R. Goroff, "An Experiment on Heat Transfer by Overstable and Ordinary Convection," *Proc. Roy. Soc. London Ser. A*, **254**, 537 (1960).
14. D. T. J. Hurle and E. Jakeman, "Significance of the Soret Effect in the Rayleigh-Jeffreys Problem," *Phys. Fluids*, **12**, 2704 (1969).
15. D. T. J. Hurle and E. Jakeman, "Soret-driven Thermosolutal Convection," *J. Fluid Mech.*, **47**, 667 (1971).
16. D. T. J. Hurle and E. Jakeman, "Natural Oscillations in Heated Fluid Layers," *Phys. Lett.*, **43A**, 127 (1973).
17. D. D. Joseph, "Uniqueness Criteria for the Conduction-Diffusion Solution of the Boussinesq Equations," *Arch. Rat. Mech. Anal.*, **35**, 169 (1969).
18. D. D. Joseph, "Global Stability of the Conduction-Diffusion Solution," *Arch. Rat. Mech. Anal.*, **36**, 285 (1970).
19. L. Koschmieder, *Advances in Chemical Physics*, Vol. 26, Wiley-Interscience, New York, 1973.

20. R. B. Lambert and J. W. Demenkow, "On Vertical Transport due to Fingers in Double Diffusive Convection," *J. Fluid Mech.*, **54,** 627 (1971).
21. J. LaSalle and S. Lefschetz, *Stability by Liapounov's Direct Method*, Academic Press, New York, 1961.
22. J. C. Legros, W. A. VanHook, and G. Thomaes, "Convection and Thermal Diffusion in a Solution Heated from Below," *Chem. Phys. Lett.*, **1,** 696 (1968).
23. J. C. Legros, W. A. VanHook, and G. Thomaes, "Convection and Thermal Diffusion in a Solution Heated from Below II: The System $CHBr_2 \cdot CHBr_2$—$CHCl_2 \cdot CHCl_2$," *Chem. Phys. Lett.*, **2,** 249 (1968).
24. J. C. Legros, W. A. VanHook, and G. Thomaes, "Thermal Diffusion in $CHBr_2 \cdot CHBr_2$—$CHCl_2 \cdot CHCl_2$ Solutions," *Chem. Phys. Lett.*, **2,** 251 (1968).
25. J. C. Legros, D. Rasse, and G. Thomaes, "Convection and Thermal Diffusion in a Solution Heated from Below," *Chem. Phys. Lett.*, **4,** 632 (1970).
26. J. C. Legros, J. K. Platten, and P. G. Poty, "Stability of a Two-Component Fluid Layer Heated from Below," *Phys. Fluids*, **15,** 1383 (1972).
27. M. J. Lighthill, "Dynamics of a Dissociating Gas; Part 2. Quasi-equilibrium Transfer Theory," *J. Fluid Mech.*, **8,** 161 (1960).
28. L. Miller and E. A. Mason, "Oscillating Instabilities in Multicomponent Diffusion," *Phys. Fluids*, **9,** 711 (1966).
29. L. Miller, T. H. Spurling, and E. A. Mason, "Instabilities in Ternary Diffusion," *Phys. Fluids*, **10,** 1809 (1967).
30. D. A. Nield, "The Thermohaline Rayleigh-Jeffreys Problem," *J. Fluid Mech.*, **29,** 545 (1967).
31. G. Oster, "Density Gradients," *Sci. Am.*, **213,** 70–76 (1965).
32. J. K. Platten and G. Chavepeyer, "Oscillations in a Water-Ethanol Liquid Heated from Below," *Phys. Lett.*, **40,** 287 (1972).
33. J. K. Platten and G. Chavepeyer, "Soret Driven Instability," *Phys. Fluids*, **15,** 1555 (1972).
34. J. K. Platten and G. Chavepeyer, "Oscillatory Motion in a Benard Cell Due to the Soret Effect," *J. Fluid Mech.*, **60,** 305, 1973.
35. Lord Rayleigh, "On Convection Currents in a Horizontal Layer of Fluid When the Higher Temperature is on the Underside," *Phil. Mag.*, **32,** 529 (1916).
36. W. H. Reid and D. L. Harris, "Some Further Results on the Benard Problem," *Phys. Fluids*, **1,** 102 (1958).
37. R. L. Sani, "On the Non Existence of Subcritical Instability in Fluid Layers Heated from Below," *J. Fluid Mech.*, **20,** 353 (1964).
38. R. L. Sani, "On Finite Amplitude Roll Cell Disturbances in a Fluid Subjected to Heat and Mass Transfer," *AIChE J.*, **11,** 971 (1965).
39. R. S. Schechter, *The Variational Method in Engineering*, Chap. 6, McGraw-Hill, New York, 1967.
40. R. S. Schechter, I. Prigogine, and J. R. Hamm, "Thermal Diffusion and Convective Stability," *Phys. Fluids*, **15,** 379 (1972).
41. R. J. Schmidt and S. W. Milverton, "On the Instability of a Fluid when Heated from Below," *Proc. Roy. Soc. London Ser. A*, **152,** 589 (1935).
42. J. Serrin, [1959]; "On the Stability of Viscous Fluid Motions," *Arch. Rat. Mech. Anal.*, **3,** 1 (1959).
43. C. C. Shir and D. D. Joseph, "Convective Instability in a Temperature and Concentration Field," *Arch. Rat. Mech. Anal.*, **30,** 38 (1968).
44. T. G. L. Shirtcliffe, "Thermosolutal Convection: Observation of an Overstable Mode," *Nature*, **213,** 489 (1967).

45. T. G. L. Shirtcliffe, "An Experimental Investigation of Thermosolutal Convection at Marginal Stability," *J. Fluid Mech.*, **35,** 677 (1969).
46. T. G. L. Shirtcliffe, "The Development of Layered Thermosolutal Convection," *Intern. J. Heat Mass Transfer*, **12,** 212 (1969).
47. T. G. L. Shirtcliffe and J. S. Turner, "Observations of the Cell Structure of Salt Fingers," *J. Fluid Mech.*, **41,** 707 (1970).
48. E. M. Sparrow, R. J. Goldstein, and V. K. Jonsson, "Thermal Instability in a Horizontal Fluid Layer: Effect of Boundary Conditions and Non-linear Temperature Profile," *J. Fluid Mech.*, **18,** 513 (1964).
49. E. A. Spiegel, "Convection in Stars: II Special Effects," in *Annual Review of Astronomy and Astrophysics*, Vol. 10, 1972, p. 269. Annual Reviews, Palo Alto, Calif.
50. M. E. Stern, "The Salt-Fountain and Thermohaline Convection," *Tellus*, **12,** 172 (1960).
51. M. E. Stern, "Collective Instability of Salt Fingers," *J. Fluid Mech.*, **35,** 209 (1969).
52. M. E. Stern and J. S. Turner, "Salt Fingers and Convecting Layers," *Deep Sea Res.*, **16,** 497 (1969).
53. H. Stommel, A. B. Arons, and D. Blanchard, "An Ocean Curiousity: The Perpetual Salt Fountain," *Deep Sea Res.*, **3,** 152 (1956).
54. M. J. Story and J. R. C. Turner, "Flow-Cell Studies of Thermal Diffusion in Liquids," *Trans Faraday Soc.*, **65,** 1523 (1969).
55. R. I. Tait and M. R. Howe, "Some Observations of Thermohaline Stratification in the Deep Ocean," *Deep Sea Res.*, **15,** 275 (1968).
56. G. Thomaes, "Recherches Sur La Thermodiffusion en Phase Liquide," *Physica*, **XVII,** 885 (1951).
57. L. J. Tichacek, W. S. Kmak, and H. G. Drickamer, "Thermal Diffusion in Liquids; The Effect of Non-Ideality and Association," *J. Phys. Chem.*, **60,** 660 (1956).
58. J. S. Turner, "Salt Fingers Across a Density Interface," *Deep Sea Res.*, **14,** 599 (1967).
59. J. S. Turner, "The Behavior of a Stable Salinity Gradient Heated from Below," *J. Fluid Mech.*, **33,** 183 (1968).
60. H. J. V. Tyrrell, J. C. Firth, and M. Zaman, "Thermal Diffusion in Solutions of Carbon Tetrachloride in Benzene," *Proc. Chem. Soc.*, **1961,** 201.
61. H. J. V. Tyrrell, J. C. Firth, and M. Zaman, "Thermal Diffusion Factors in the Benzene-Methanol System," *Proc. Chem. Soc.*, **1965,** 3613.
62. M. G. Velarde and R. S. Schechter, "Thermal Diffusion and Convective Stability (III): A Critical Survey of Soret Coefficient Measurements," *Chem. Phys. Lett.*, **12,** 312 (1971).
63. M. G. Velarde and R. S. Schechter, "Thermal Diffusion and Convective Stability. II: An Analysis of the Convected Fluxes": *Phys. Fluids*, **155,** 1707 (1972).
64. G. Veronis, "On Finite Amplitude Instability in Thermohaline Convection," *J. Marine Res.*, **23,** 1 (1965).
65. G. Veronis, "Effect of a Stabilizing Gradient of Solute on Thermal Convection," *J. Fluid Mech.*, **34,** 315 (1968).
66. G. Walin, "Note on the Stability of Water Stratified by both Heat and Salt," *Tellus*, **16,** 3 (1964).
67. S. Whitaker and R. L. Pigford, "Thermal Diffusion in Liquids," *Ind. Eng. Chem.*, **50,** 1026 (1958).
68. D. J. Wollkind and H. L. Frisch, "Chemical Instabilities: I. A Heated Horizontal Layer of Dissociating Fluid," *Phys. Fluids*, **14,** 13 (1971).
69. W. Weinberger, "The Physics of the Solar Pond," *Solar Energy*, **8,** 45 (1964).

AUTHOR INDEX

Numbers in parentheses are reference numbers and show that an author's work is referred to although his name is not mentioned in the text. Numbers in *italics* indicate the pages on which the full references appear.

Aarseth, S. J., 117(17), *137, 144*
Acrivos, A., *193, 194, 197, 204, 211*
Agar, J. N., *282, 298, 299*
Ahlers, G., *209, 212*
Ailawadi, N. K., 4(40), 58(40), 60(40), *103*
Albada, T. S., v., *138, 144*
Alder, B. J., 2(22), *8,* 21(22), 23(22), *51,* 52(22), *53,* 55(22), *77, 102, 105, 106, 117, 144*
Alexander, B., 163(73), *176*
Allin, E. J., 5(53), *104*
Anderson, H. C., 4(39), 58(39), 60(39), *103*
Antonov, V. A., 115(9), 132(9), *144*
Armstrong, D., *167, 176*
Arnau, C., 245(46), 246(46), 251(73), *261*
Arons, A. B., 274(53), *301*
Ash, B. J., *152, 175*
Avsec, D., *183, 188, 206, 210, 212*

Bagus, P. S., *239, 246,* 250(71), *260, 261*
Baier, R. E., 163(68), *176*
Baines, P. G., *275, 277, 278, 299*
Baird, N. C., 255(100), *262*
Bang, N. U., *160, 175*
Bannister, T. C., *194, 211*
Barlow, G. H., 163(70), *176*
Barrow, P., 165(84), *176*
Barton, P. G., 153(39), *175*
Basch, H., 245(46), 246(46), 249(69), 251 (75), 253(96), *261, 262*
Baumann, G., *133, 144*
Bdzil, J., *294, 299*
Beck, E. A., 161(62), *175*
Becker, E. L., *169, 170, 176*
Belton, P. S., *290, 299*
Bénard, H., *177, 187, 188, 210*
Bender, C. F., 240(34), 248(67), *260, 261*
Benedek, G. B., 2(25), 27(25), 88(120), 89(120a), *102, 106*
Ben-Reuven, A., 4(33,36,38), 19(33), *57,* 58(38), *60, 93, 103, 106*
Berg, J. C., *193, 211*
Berne, B. J., 4(40), *8,* 21(70), 23(70), *42,* 59(40,99), 60(40), *103, 104*
Berthier, G., 248(65), 257(114), *261, 263*
Bertoncini, P., 248(68), *261*
Beveridge, D. L., 255(99,103), *262, 263*
Biggs, B. J., *152–154,* 155–157(42), *175*
Billingsley, F. P., *257, 263*
Bird, R. B., 74(112), 88(121), *106*
Birnbaum, G., 2(3,14), *3,* (53,55), *7, 19, 37, 39, 102, 104, 105*
Birss, F. W., *239, 260*
Bishop, M., *8,* 21(70), 23(70), *42, 104*
Blanchard, D., 274(53), *301*
Block, M. J., *192, 211*
Bloembergen, N., 5(50), *103*
Blömbäck, B., 158(53), 159(53), *160, 161, 175*
Blömbäck, 158(53), 159(53), 161(53,63), *175*
Bloor, J. E., 257(117), *263*
Blum, L., 5(45), *103*
Bodar, N., 256(107), *263*
Boissonas, R. A., 168(94), *176*
Born, G. V. R., 156(45), *175*
Born, M., *12,* 25–27(86), *32, 61,* 62(86), *104, 105, 239, 259*
Boys, S. F., *239,* 245(29), *260*
Brand, E., *148, 174*
Brash, J. L., 163(66), *176*
Brillouin, L., *248, 261*
Broida, H. P., 4(31), *103*
Brown, R. D., 150(19), *174*
Brown, W. F., Jr., 24(84), 27(84), *105*
Bucaro, J. A., 2–4(6,11), 7(11,69), 39(6, 69), *40, 41,* 43(69), *102, 104*
Buckingham, A. D., 6(75), 7(67), *49, 50, 61,* 62(104), *63,* 67(67), *104, 105*
Buenker, R. J., 254(89,94), *262*
Bullough, R. K., *6, 30, 104*
Bunker, R. J., 251(78), *262*
Busse, F. H., *191, 195, 291, 203–206, 210, 211, 212*

Cabannes, J., 3(26), *102*
Cade, P. E., 239(27), 244(27), *260*
Caldwell, D. R., *291, 293, 299*
Candau, S. J., 4(32), *103*
Carini, M., 163(66), *176*
Carr, H. Y., 88(120), *106*
Catton, I., *196, 211*
Chaikin, S. W., 163(66), *176*
Chandra, K., *188, 200, 206, 210*
Chandrasekhar, S., 111(1), *112,* 113(1), *131,* 141(1), *143, 144, 177, 183, 201, 207, 210, 268, 299*
Charlson, G. S., *194, 196, 197, 211*
Chavepeyer, G., 266(32,34), *287, 288, 291–293, 300*
Chen, M. M., *200, 206, 211*
Chorin, A. J., *191, 210*
Christoffersen, R. E., 245(48), 247(60), *252,* 253(83), 254(79,93), *261, 262*
Christopherson, D. G., *181, 210*
Chu, B., 2(1), 16(1), 27(1), *101*
Chung, C. H., 4(39), 58(39), 60(39), *103*
Clementi, E., 245(44,49), *246,* 253(80,81), 254(88,90), *261, 262*
Cochrane, C. G., 165(85), 166(85), *176*
Cohen, E. R., *39, 105*
Cohen, M., 248(66), *261*
Cole, R. H., 65(106), *74, 76,* 77(106), *105*
Coln, M., 150(21), *174*
Connolly, J. W. D., 258(118), 259(118), *263*
Corey, R. B., 148(13), *174*
Cotton, F. A., 258(124), *263*
Courtens, E., 4(36), 60(36), *103*
Crawford, M. F., 5(52), *104*
Crawford, R. K., 77(115), *106*
Csizmadia, I. G., 245(44), *261*
Currie, J. G., *200, 211*
Curtiss, C. F., 74(112), 88(121), *106*

Dalgarno, A., 248(66), *261*
Danchin, A., 150(23), *174*
Daniels, W. B., 2–4(12), 77(115), *102, 106*
Darwin, C. G., 26(87), *32, 105*
Das, G., *239,* 247(23), 248(68), 253(23), *260, 261*
Dauzere, C., *187, 200, 210*
Davidson, A. F., 111(4), *143*
Davidson, E. R., *234,* 248(67), 251(74), *260, 261, 262*
Davie, W. E., *151–154, 175*
Davies-Jones, R. P., *196, 211*
Davis, D. R., 245(44), *261*
Davis, S. H., *188, 196, 202, 204, 205, 210, 211, 212*
Davydov, A. S., 239(1), *259*
Dawes, E. A., 149(16), 171(16), *174*
Dayhoff, M. J., 148(8), *160,* 162(8), *174*
Deardorff, J. W., *191, 197, 198, 200, 202, 207, 210, 211*
Del Bene, J., 254(95), 256(108), *262, 263*
Demenkow, J. W., *280, 281, 300*
Demuynek, J., 254(92), *262*
Denson, K. W. E., *152, 175*
Denys, J., *164, 176*
Deutch, J. M., 5(43), *103*
Dewar, M. J. S., 255(100,101,104), 256 (107), *262, 263*
Dierksen, G. H. F., 254(86,87), *262*
Di Prima, R. C., *202, 203, 206, 212*
Ditchfield, R., 245(43), 254(95), *261, 262*
Dobosh, P. A., 255(99), *262*
Doolittle, R. F., *161, 175*
Doremus, J.–P., *133, 144*
Dougherty, T. S., *204, 212*
Drickamer, H. G., 290(57), *301*
Dropkin, D., *191, 211*
Dry, R. M. L., *167, 176*
Dunmur, D. A., 7(67), *49,* 65(67), *104*
Dunning, T. H., 239(16), *246,* 251(75), *260, 261, 262*
Du Pre, D. B., 2–4(9,12), *6,* 40(9), *49, 102*
Dutton, R. C., 163(68), *176*

Eck, R. V., 148(8), *160,* 162(8), *174*
Eckart, C. E., 239(2), *259*
Eckhaus, W., *203, 212*
Ehoyn, D., 163(70), *176*
Einstein, A., *2, 6, 102*
Elbert, D. D., 131(21), *144*
Elbert, S. T., 251(75), *262*
Ellenson, W. D., 2–4(8), 7(8), *8, 38, 47, 55, 64,* 65(8), *102*
Ellingsen, T., *204, 212*
Elliott, D. F., 168(93), *176*
Enderby, J. E., 28(89), *105*
Engle, A., 163(73), *176*
Esnouf, M. P., *153, 175*

Fabelinskii, I. L., 2(1), 4(35), 16(1), 27(1),

60(35), *101, 103*
Felix, M. R., *133, 144*
Ferry, J. D., *160, 175*
Fink, W. H., 240(36), *261*
Firth, J. C., 290(60,61), *301*
Fischer, H., 256(106), *263*
Fisher, I. Z., *15, 104*
Fixman, M., *6, 30, 104*
Fleury, P. A., 2–4(9,12), *6,* 40(9), *49, 102*
Fock, V., *239, 259*
Fonsen, T., 156(44), *175*
Forgenson, L., 157(47), *175*
Forster, D., 4(40), 58(40), 60(40), *103*
Foster, T. C., *199, 211*
Fraga, S., *239, 260*
Franken, P. A., 5(48), *103*
Frenkel, J., *239, 260*
Frenzen, P., *207, 212*
Frey, E. K., *167, 176*
Frisch, H. L., 5(43), *6, 34,* 91(62), *103, 104, 293, 294, 299, 301*
Fromm, J. E., *191, 211*
Frost, A. A., 247(59), *252, 261*
Fucker, M. L., 163(67), *176*

Gage, K. S., *207, 212*
Gallagher, A. P., *207, 212*
Gardner, R. L., 164(75), *176*
Gebbie, H. A., 5(104), *104*
Gelbart, W. M., 2(21), 5(51), *8,* 14(73), 16(79), 18(21), 19(21), 21(21), 23(21), 24(51), 28(73), 39(21), 44(21), 46(21), 68(109), 79(117), 85(117), 92(73), *102, 103, 104, 105, 106*
Geller, M., 245(44), *261*
Gennes, P. G., d., 49(94), *105*
Genson, D. W., 252(79), 254(79), *262*
Gershon, N. D., 4(33,38), 19(38), *57,* 58 (38), 60(33,38), *103*
Gersten, J. I., 2(5,19), 3(5), 4(5), 7(19), 19(19), *40, 102*
Giglio, M., 88(120), 89(120a), *106*
Gilbert, T. L., *239,* 246(51), 253(25), *260, 261*
Gill, A. E., *275, 277, 278, 299*
Gille, J., *191, 211*
Gingrich, N. S., 50(95), *105*
Gjevik, B., *204, 212*
Glansdorff, P., *267, 299*
Glauber, R. J., 28(89), *105*
Glynn, M. F., 157(47), *175*
Goddard, W. A., 239(16), 251(75), 253 (82), *260, 262*
Goldberg, B., 165(84), *176*
Goldfinger, S. E., 169(96), *176*
Goldstein, A. W., *200, 211*
Goldstein, R. J., *209, 212,* 274(48), *301*
Goldwater, W. H., 148(1), *174*
Gordon, R. G., 4(30), *103*
Gorkov, L. P., *201, 211*
Gornall, W. S., 2–4(10), *41, 102*
Goroff, I. R., *277, 299*
Graham, A., *187, 203, 206, 210*
Gray, C. G., 2(17), *8,* 21(17), *46, 102, 105*
Green, H., 165(84), *176*
Green, H. S., 24(84), 27(84), *105*
Gregg, M., 266(10), *299*
Grodzka, P. G., *194, 211*
Groot, S. R., d., 282(11), 283(11), *299*
Guarder, A., 156(44), *175*
Gueron, M., 150(23), *174*
Gutschick, V. P., *22, 55,* 78(80), *104*
Guttmann, S., 168(94), *176*

Haanen, C., 153(40), *175*
Hall, C. E., 158(48), *175*
Hall, G. G., *239, 260*
Hall, L. H., 2–4(8), 7(8), *8,* 21(8), *38, 47, 55, 64,* 65(8), *102*
Hamberg, V., 163(72), *176*
Hamm, J. R., *285,* 288(40), *300*
Hanahan, P. J., 153(39), *175*
Handy, N. C., 247(57), *261*
Harget, A., 256(107), *263*
Hariharan, P. C., 246(54), *261*
Harris, D. L., *180, 210,* 274(36), *300*
Harris, F. E., 244(41), 245(43), *260, 261*
Harris, R. A., *93, 106*
Harrison, M. C., 245(44), *261*
Hartree, D. R., *239, 259*
Haselbach, E., 256(107), *263*
Hazi, A., 240(35), *260*
Hedin, L., 247(62), 253(97), *261, 262*
Hehre, W. J., 245(47), 249(69), 254(84,85), *261, 262*
Hellem, A., 156(44), *175*
Heller, D. F., 68(109), *105*
Hellwarth, R. W., 2(15), *61, 93, 102*
Hemker, H. C., 151(27), 171(27), 172(27), *173, 175*

Hemker, P. W., 151(27), 171(27), 172(27), *173, 175*
Henderson, D., 66(107), *105*
Herring, J. R., *202, 212*
Hill, T. L., 15(78), 91(78), *104*
Hinkley, R. K., 253(96), *262*
Hinze, J., *239,* 247(24), 253(24), *260*
Hirschfelder, J., 74(112), 88(121), *106*
Ho, J. H. K., 2–4(13), 8(13), *41, 102*
Hoard, C. O., *194, 204, 211*
Hodge, R. L., 168(87), *176*
Hoerner, S., v., *117, 144*
Hoffmann, R., 257(110,112), *263*
Homsy, G. M., *208, 212*
Hopf, H., 97(22), 98(22), 100(22), 102 (22), *103,* 109(22), 116(22), 138(22), *169*
Horton, E. W., 168(93), *176*
Hosteny, R. P., 240(34), *260*
Hovig, T., 157(47), *175*
Howard-Lock, H. E., 2–4(10), 41(10),*102*
Howe, M. R., *280, 301*
Huang, K., 12(72), *104,* 239(1), *259*
Hudson, J. L., *208, 212*
Hunt, W. J., *239,* 251(75), *260, 262*
Hurle, D. T., *199, 211, 285, 288, 293, 299*
Huzinaga, S., *239,* 245(46), *246,* 251(73), *260, 261, 262*
Hylleraas, E. A., *239, 259*
Hynes, J. T., 5(43), *103*

Jackson, J. D., *106*
Jaffé, H. H., 256(108), *263*
Jakeman, E., *199, 211, 285, 288, 293, 299*
Jansen, L., 27(88), 34(88), *78, 105, 106*
Jaquenoud, P. A., 168(94), *176*
Jeffreys, H., *193, 211*
Jirgensons, B., *148, 174*
Jobin, F., 153(38), *175*
Johnson, K. H., 258(118,119,123,124), 259(126), *263*
Johnson, S. A., *156, 157, 175*
Jones, R. W., 150(24), *174*
Jonsson, V. K., 274(48), *301*
Joseph, D. D., *197, 211, 295, 296, 299*

Kadanoff, L. P., 4(28), *102*
Kahn, L. R., 253(82), *262*
Kammer, W. E., 251(78), *262*
Kapral, R., 5(46), *103*
Kassell, W. B., 148(1), *174*
Kay, K. G., 244(41), *260*
Keele, C. E., *167, 168, 176*
Kelley, H. P., 247(61), *261*
Kendrew, J. C., *148, 174*
Kerr, J., *61, 105*
Kerwin, D. J. G., 2(16), 7(16), 21(16), 35 (16), *102*
Keyes, T., 4(39), *58, 59, 60, 103, 105*
Kezdy, F. J., *153, 175*
Khinchin, A. I., 127(20), 128(20), *144*
Kielich, S., 6(58), *23, 104*
Kingdon, H. S., *152, 175*
Kirdyashkin, A. G., *191, 200, 211*
Kirkwood, J. G., *24, 27, 93, 105*
Kirschenbaum, D. M., 148(3,4), *174*
Kistenmacher, H., 245(49), 254(88), *261, 262*
Kivelson, D., 4(39), *58, 59,* 60(39), *103, 105*
Klein, K. G., 163(66), *176*
Klopman, G., 255(101,102), *262*
Kmak, W. S., 290(57), *301*
Knirk, D. L., 5(45), *103*
Kobe, D. H., 249(64), *261*
Koenig, S. H., 150(19), *174*
Kogelman, S., *202, 203, 206, 212*
Kohn, W., 258(120), *263*
Kollmar, H., 256(106), *263*
Komarov, L. I., *15, 104*
Koopmans, T. A., *249, 261*
Koppel, G., *158, 175*
Korringa, J., 258(120), *263*
Koschmieder, E. L., *183, 193–195, 200, 202, 207, 208, 211, 212, 268, 299*
Koshland, E. D., 150(18), *174*
Kowalski, E., 164(78), *176*
Kraemer, W. P., 254(86,87), *262*
Kranendonk, J., v., 2(16), 7, 21(6), 24(82), *35, 102, 104*
Krauss, M., 253(96), *262*
Kraut, H., *167, 176*
Krishnamurti, R., *197, 200, 204, 211, 212*
Kuebler, N. A., 245(46), 246(46), *261*
Kulacki, F. A., *209, 212*
Kuo, H. L., *202, 211*
Küppers, G., *207, 212*
Kurth, R., *114, 144*

Lachmann, P. J., 165(81,82), *176*

Laki, K., *147, 152, 175*
Lallemand, P., 2(4,18), 3(4), 4(4), 7(8), 19 (18), *38, 49,* 80(4), *102*
Lambert, R. B., *280, 281, 300*
Land, R. H., 244(42), *261*
Landau, L. D., *139, 144*
Langevin, P., *61, 105*
Langhoff, S. R., 251(74), *262*
Larson, R. B., *133, 144*
Larsson, K. E., 28(90a), *49, 105*
LaSalle, J., *295, 300*
Lathan, W. A., 254(84,85), *262*
Lee, E. P., 111(2), 112(6), 131(21), *143, 144*
Lefebvre, R., *239, 260*
Lefschetz, S., *295, 300*
Legros, J. C., *285, 290, 291, 300*
Lighthill, M. J., *293, 300*
Lehn, J. M., 254(91), *262*
Leland, S., 156(44), *175*
Leontiev, H., *191, 200, 211*
Levine, 151(26), *175*
Levine, H. B., 2(14), *3,* 5(55), *7,* 19(14), *37,* 39(14b), *69, 71,* 78(111), *102, 104, 106*
Levy, B., 248(65), *261*
Lewis, G. P., 168(93), *176*
Lewis, J. C., 2(16), *7,* 21(16), *102*
Liang, S. F., *194, 197, 211*
Lifshitz, E. M., *139, 144*
Lipps, F. B., *191, 207, 210, 212*
Litovitz, T. A., 2(6,11), 3(6,11), 4(6,11, 32), *7,* 39(6,69), *40, 41,* 43(69), *102, 103, 104*
Liu, B., 246(54), *261*
Lloyd, R. V., 183(49), 184(51), *201*
Locke, J. L., 5(52), *104*
Loehr, T. M., *105*
Loewy, A. G., *158, 175*
Lorand, L., *153, 158, 159,* 160(52), *175*
Lortz, D., *191, 195, 201, 207, 211, 212*
Low, A. R., *180, 210*
Lowdin, P. O., *239, 260*
Lowe, R. D., 166(87), *176*
Lundqvist, S., 247(62), 253(97), *261, 262*
Lyman, D. J., 163(66), *176*
Lynden-Bell, D., 111(2,3), 115(10,13), 132(10), *133, 143, 144*
MacDonald, J. K. L., *239, 259*
MacFarlane, R. G., *151–155,* 164(74), *175, 176*
McGinn, G., 242(37), *260*
McIntyre, D., 2(2), 4(2), 16(2), 27(2), 92 (2), *101*
McKendrick, A., *172, 176*
McKenna, J., *6, 34, 91, 104*
McKoy, V., *22, 55,* 67(108), 78(80), *104, 105,* 251(75,76,77), *262*
McLean, A. D., 244(40), 246(54), *260, 261*
McNichol, G. P., 163(69), 164(69), *176*
McQuarne, D. A., 5(55), *69, 71,* 78(111), *104, 106*
McTague, J. P., 2(3,8,9), 3(3,8,9), 4(8,9), *6–8, 19,* 21(8), *22, 38, 40, 47, 49, 55, 58, 59, 64,* 65(8), 67(108), 78(80), *102,* 104, 105
McWeeny, R., *239, 260*
Maggiora, G. M., 245(48), *261*
Magnussom, St., 161(64), 163(64), *175*
Malkus, W. V. R., *201, 211*
Mandel, M., 27(88), 34(88), *105*
Marboix, H., d., *164, 176*
Markham, J. W., *167, 176*
Marron, M. T., 247(57), *261*
Marshall, A. G., 5(47), *103*
Martin, P. C., 4(28), *102*
Martin, R. S., 240(32), *260*
Mason, E. A., 281(28,29), *300*
Mason, R. G., 161(65), *176*
Matcha, R., 246(52), *261*
Matze, R., 153(48), *175*
May, A. D., 2(16), 5(53), 7(16), 21(16), 35(16), *102, 104*
Mazur, P., *6, 27, 30, 34,* 62(105), *78, 104, 105, 106,* 282(11), 283(11), *299*
Mehl, J., 253(81,90), *262*
Melmon, K. L., 169(96), *176*
Mercer, A. McD., *207, 212*
Merrache, D., 161(61), *175*
Merzbacher, E., 244(39), *260*
Michejda, C. J., 254(89), *262*
Michels, H. H., 244(41), *260*
Mihaljan, J. M., *179, 210*
Mihalyi, E., 158(49), *175*
Mikolaj, P. G., 50(95), *105*
Miles, A., 167(89), *168, 176*
Miller, K. D., *153, 175*
Miller, L., *281, 300*

Miller, M., 160(157), *175*
Miller, R. H., 117(14), 124(18), 126(19), 133(18,19), 134(24), 138(14), *144*
Millie, P., 257(114), *263*
Milverton, S. W., *189, 210, 290, 300*
Moller, C., 248(66), *261*
Mollison, P. L., 165(80), *176*
Montroll, E. W., 173(99,100), 174(100, 101), *176*
Moore, S., 148(5,6), *174*
Mori, H., *59, 105*
Morrison, P. R., 160(56), *175*
Moskowitz, J. W., 245(44), *261*
Mountain, R. D., 4(27), *6,* 21(64), *34, 84, 90, 91, 102, 104*
Mross, G. A., 161(59), *175*
Müller, U., *196, 207, 211, 212*
Muller-Eberhard, H. J., 165(83,85), 166 (85), *176*
Mulliken, R. S., *257, 263*
Mustard, J. F., *157, 175*

Nakagawa, Y., *207, 212*
Navon, G., 150(20), *174*
Nelkin, M., 4(29), *103*
Nesbet, R. K., *240, 260*
Neumann, J., v., 112(5), *144*
Newton, M. D., 257(115,116), *263*
Nicholson, B. J., 257(115), *263*
Niessen, P., *207, 212*
Niessen, W., v., 253(81), 254(87,90), *262*
Nield, D. A., *193, 194, 199, 211, 273,* 274(30), *279, 300*
Niewiarowski, S., 164(78), *176*
Nijboer, B. R. A., 74(113), *106*
Nishizawa, E., 157(47), *175*
Norman, J. G., 258(118,124), *263*
Nussenzweig, V., 164(79), *176*

O'Brien, E. F., *22, 55, 78, 104*
Oden, L., 66(107), *105*
Ogura, Y., *203, 212*
Oiann, H., *203, 212*
Oksengorn, B., 2(7), 4(7), 35(7), 49(7), 50(7), *102*
O'Leary, B., 255(102), *262*
Oppenheimer, J. R., *239, 259*
Orcutt, R. H., 65(106), *74, 76,* 77(106), *105*
Ortoleva, P. J., 4(29), *103*
Oster, G., *281, 300*
Ostlund, N. S., 257(116), *263*
Ostrach, S., *178, 210*
Ostriker, J. P., 111(4), *143*
Ostrowski, G., 28(90b), 49(90), *105*
Owren, P. A., *152,* 156(44), *175*
Oxburgh, E. R., *178, 210*
Oxtoby, D., 14(73), 28(73), 79(117), 85 (117), 92(73), *104, 106*

Paalamn, H. H., 15(77), *104*
Packham, M. A., 157(47), *175*
Paladini, A. C., 166(88), *176*
Pallas, S. G., *190, 191, 195, 201, 210*
Palm, E., *203, 204, 212*
Palmer, H. J., *193, 211*
Pasmanter, R. A., *96, 106*
Paul, R., 6(59), *104*
Pauling, L., 148(13), *174*
Peacocke, A. R., 150(22), *174*
Pearson, J. R., *192–194, 211*
Pechet, L., 163(73), *176*
Peart, W. S., 166(86), 167(86), *176*
Pecora, R., 4(30,39), 5(43,47), *15,* 57 (30a), 58(39), 60(39), *103, 104*
Pellew, A., *180, 210*
Percus, J. K., *44, 105*
Perutz, M. F., *148, 174*
Peticolas, W. L., 5(49), *103*
Peyerimhoff, S. D., 251(78), 254(89,94), *262*
Pierce, J. V., 168(95), *176*
Pigford, R. L., *290, 301*
Pike, E. R., 199(69), *211*
Pilar, F. L., 239(1), *259*
Pilling, J., 164(74), *176*
Pincus, P. A., *105*
Pings, C. J., 15(77), 50(95), *104, 105*
Pinnow, D. A., 4(52), *103*
Pipano, A., 240(34), *260*
Platten, J. K., 266(32,34), *285, 287, 288, 291–293, 300*
Platzman, G. W., *202, 211, 213*
Plesset, M. S., 248(66), *261*
Plows, W. H., *191, 201, 210*
Polezhaev, V. T., *191, 210*
Poll, J. D., 24(82), *104*
Popkie, H., 245(49), 254(88), *261, 262*
Pople, J. A., *61, 63, 105,* 245(47), 246(54), 249 (69), 254(84,85,95), 255,(98,99.103), 256 (98,105), 257(117), *261, 262, 263*

Postma, B. J., 60(105), *105*
Poty, P. G., *285, 300*
Preuss, H., *239,* 247(58), *260, 261*
Prigogine, I., 112(7), *126, 132, 144, 267, 285,* 288(40), *299, 300*
Proceedings of the International Colloquim of the CNRS on the Scattering of the Light Fluids (July 15–17, 1971, Paris), 4(41), 58(41), 60(41), 61(41), *103*
Pullman, A., 255(102), *262*

Quick, A. J., *152, 175*

Radom, L., 249(69), 254(84,85), *261, 262*
Rahman, A., *8,* 21(70), 23(70), *42, 104*
Ralph, H. I., 2(17), *8, 46, 102*
Ramshaw, J. D., *24, 104*
Randolph, P. D., 28(90b), 49(90), *105*
Ranganathan, S., *102*
Ransil, B. J., 247(56), *261*
Ratnoff, O. D., *151–154, 175*
Rasse, D., 291(25), *300*
Rayleigh, Lord, *177, 178, 210, 274, 300*
Reid, W. H., 180, 207, 210, 212, 274 (36), *300*
Reiner, J. M., 149(17), 171(17), *174*
Reinsch, C., 240(31), *260*
Reuben, J., 150(21), *174*
Rice, S. A., 240(35), *260*
Richards, R. E., 150(22), *174*
Richards, W. G., 253(96), *262*
Robbins, K. C., *163, 176*
Roberts, P. H., *203, 204, 212*
Roberts, P. M., *209, 212*
Robertson, C. R., *194, 204, 211*
Robin, M. B., 245(46), 246(46), *261*
Rodman, N. F., 161(65), *176*
Roothaan, C. C. J., *239,* 244(38,27), 246(51), *260, 261*
Rose, J., 251(76,77), *262*
Rosenberg, A., 5(53), *104*
Rosenfeld, L., 24(85), *27, 105*
Roskam, J., 156(43), *175*
Rossby, H. T., *190, 197, 208, 210*
Rostoker, N., 258(120), *263*
Rouse, R. A., 247(59), 252(59), *261*
Rowe, D. J., 251(77), *262*
Rowe, J. M., 28(90b), 49(50), *105*
Rowrill, H. C., 157(147), *175*
Ruedenberg, K., *239,* 244(28), 246(50), *257, 260, 261, 263*
Ryan, F. Y., 148(1), *174*

Sabin, J. R., 259(125), *263*
Sachs, L. M., 244(38), 245(44), *260, 261*
Saidel, L. J., 148(1), *174*
Sakurai, J. J., 10(71), 11(71), *104*
Salmon, L., 239(28), 244(28), *260*
Salsburg, Z. W., 5(45), *103*
Samson, R., *93, 106*
Sanger, F., *148, 174*
Sani, R., d., *194, 196, 197, 211, 278, 295, 297, 298, 300*
Santry, D. P., 255(98), 256(98), *262*
Saslow, W. C., 115(13), 138(25), *144*
Saunders, O. A., *189, 200, 210*
Scanlon, J. W., *193, 211*
Schaefer, H. F., *239,* 247(18), 250(71), *260, 261*
Schechter, R. S., *178, 210, 279, 285, 288–290, 297, 298, 300, 301*
Schluter, A., *191, 195, 201, 211*
Schmidt, P. W., 88(123), *106*
Schmidt, R. J., *189, 210, 211, 290, 300*
Schneck, P., *191, 210*
Schoenmakers, J. G. G., *153, 175*
Schurr, J. M., 5(44), *103*
Schwab, H. J., *208, 212*
Schwartz, M., 242(37), *260*
Schwiderski, E. W., *208, 209, 212*
Scornik, O. A., 166(88), *176*
Scriven, L. E., 192, 211
Sears, V. F., 7(68), 35(68), *104*
Seegers, W. H., *147, 153, 157, 158, 175*
Segal, G. A., 255(98), 256(98,105), *262, 263*
Segel, L. A., *177, 193, 197, 204, 210, 211, 212*
Seligman, M., 164(79), *176*
Sengers, J. V., 2(2), 4(2), 16(2), 27(2), 92(2), *101*
Serrin, J., *295, 300*
Severne, G., 112(7), *126, 132, 144*
Shannon, C. E., 127(20), 128(20), *144*
Shapiro, S. L., 4(31), *103*
Sharp, A. A., 164(77), *176*
Shavitt, I., *240,* 245(43), *260, 261*
Sheard, B., 150(22), *174*
Shelton, D.P., 2–4(13), 8(13), *41, 102*

Sher, D., 138(25), *144*
Shibuya, T., 251(76,77), *262*
Shih, S., 254(89,94), *262*
Shin, H. K., 2(20), *8,* 19(20), *41, 102*
Shipman, L. L., 245(48), 247(60), 252 (79), 254(79), *261, 262*
Shir, C. C., *295, 300*
Shirtcliffe, T. G. L., 266(47), *280–282, 293, 297, 300, 301*
Shuler, K., 174(101), *176*
Shulman, 160(57), *175*
Siegmiller, J. E., 169(96), *176*
Silverstone, H. J., 244(41), 247(57), 250 (72), *260, 261*
Silveston, P. L., *190, 210*
Sköld, K., *28, 49, 105*
Slater, J. C., *239, 258, 259, 263*
Slayter, H. S., 158(48), *175*
Slusher, R. E., 2-4(5), 40(5), *102*
Smith, F. C., 258(119,123), *263*
Smith, K. A., *192, 211*
Snyder, L. C., 249(69), 253(96), *261, 262*
Sobel, H. M., 148(12), *174*
Soberman, R. K., *194, 211*
Sogin, H. H., *189, 210*
Somerscales, E. F. C., *191, 204, 211, 212*
Somerville, R. C. J., *191, 197, 198, 200, 202, 207, 210, 211, 212*
Southwell, R. V., *180, 210*
Spackman, D. H., 148(5,6), *174*
Sparrow, E. M., 274(48), *301*
Spiegel, E. A., *179, 210,* 266(49), *301*
Spurling, T. H., 281(29), *300*
Starunov, V. S., 4(34,35), 60(35), *103*
Steele, W. A., 4(30), 57(30a), 60(100), *103, 105*
Stegeman, G. I. A., 4(37), 60(37), *103*
Stein, W. H., 148(5,6), *174*
Steinborn, E. O., 239(28), 244(28), *260*
Stephen, M. J., 5(47), *6, 49, 50, 103, 104*
Stern, M. E., *270, 277, 278, 280, 281, 301*
Sternling, C. V., *192, 211*
Stevens, R. M., 244(42), 246(55), *261*
Stoicheff, B. P., 2(10), 3(10), 4(10,37), 5(53), 41(10), 60(37), *102, 103, 104*
Stommel, H., *274, 301*
Stone, N. W. B., 5(53), *104*
Stork, K., *196, 211*
Story, M. J., *290, 301*
Strauss, H. L., 2(22), *8,* 21(22), 23(22), *51,* 51(22), *53,* 55(22), *77, 102, 106*
Streitwieser, A., 257(111), *265*
Stuart, J. T., *181, 200, 204, 210, 211, 212*
Stryland, J. L., 5(53), *104*
Studebaker, J., 150(19), *174*
Summaria, L., 163(70), *176*
Surko, C. M., 2-4(5), 40(5), *102*
Sutcliffe, B. T., 245(44), *261*
Switzlski, J. D., 242(37), *260*
Szasz, L., 242(37), *260*
Szöke, A., 4(36), 60(36), *103*
Szymanski, H. A., 5(42), *103*

Tabisz, G. C., 2(7,13), 3(13), 4(7,13), 8 (13), 35(7), *41,* 49(7), 50(7), *102*
Tait, R. I., *280, 301*
Tanaka, M., *6, 104*
Theimer, O., 6(59), *104*
Thibeau, M., 2(7), 4(7), *35, 49, 50, 102*
Thirlby, R., *209, 212*
Thirring, W., 115(11), 132(11), *144*
Thomaes, G., *290,* 291(25), *300, 301*
Thompson, C. W., 50(95), 88(123), *105, 106*
Thompson, H. A., *189, 210*
Tichacek, L. J., *290, 301*
Tiganov, E. V., 4(35), 60(35), *103*
Tillet, W. S., 164(75), *176*
Tippelskirch, H., v., *188, 191, 203, 210, 212*
Tritton, D. J., *208, 212*
Tuppy, H., 148(7), *174*
Turcotte, D. L., *178, 210*
Turner, J. S., 266(47), *280–282, 290, 298, 299, 301*
Tyrrell, H. J. V., *290, 299, 301*

Undheim, B., *239, 259*
Unny, T. E., *207, 212*

Vane, J. R., 166(87), *176*
Van Hook, W. A., 290(22,23), *300*
Van Hove, L., *15,* 28(74), 74(113), *104, 106*
Veillard, A., 246(52), 254(92), 257(114), *261, 262, 263*
Velard, M., *178, 210, 279, 289, 290, 298, 301*
Verhaegen, G., 253(97), *262*
Veronis, G., *277, 297, 298, 301*

Vidal, A., *194, 197, 211*
Vineyard, G. H., 28(89), *105*
Vlasyak, M. P., *191, 210*
Vodar, B., 2(7), 4(7), 35(7), 49(7), 50(7), *102*
Volterra, V., 2–4(6), 7(69), *39, 43,* 60 (39), *102, 103, 104*
Vroman, L., 163(67), *176*

Wahl, A. C., *239,* 244(27,42), 247(23), 248(68), 253(23), *260, 261*
Wahlgren, U., 259(126), *263*
Wainwright, T. E., 51(96), *105*
Walin, G., *276, 301*
Walker, T. E. H., 253(96), *262*
Wall, W. R., 2–4(13), 8(13), *41, 102*
Wallén, P., *163, 176*
Ward, J. F., 5(48), *103*
Watson, J., 148(11), *174*
Webster, M. E., 168(95), 169(96), *176*
Weeks, J. D., 240(35), *260*
Weinberg, M., 5(46), *103*
Weinberger, W., 266(69), *301*
Weiss, J. J., 2(22), *8,* 21(22), 23(22), *51,* 52(22), *53,* 55(22), 74(114), *102, 106*
Weiss, A. W., 244(38), *260*
Welsh, H. L., 5(52,53), *104*
Werle, E., *167, 176*
Whittaker, S., *290, 301*
Whitehead, J. A., *200, 206, 211*
Whitten, J. L., 245(45), *246, 261*
Wielen, R., 117(17), *144*
Wilkinson, J. H., *240, 260*
Willis, G. E., *197, 198, 200, 202, 211*
Wilson, T. M., 258(122), *263*
Wipff, G., 254(91), *262*
Wold, F., *148,* 149(15), *174*
Wolf, E., 25–27(86), *32,* 62(86), *105*
Wollkind, D. J., *293, 301*
Wood, J. H., 258(122), *263*
Wood, R., 115(10), 132(10), *144*
Woodward, R. B., 257(112), *263*
Worlock, J. M., 2–4(12), *102*

Yevick, G. J., *44, 105*
Yin, M. L., 250(72), *261*
Yip, S., 4(29,39), 58(39), 60(39), *103*
Yoshimine, M., 246(54), *261*
Yvon, J., *24, 27, 105*

Zaman, M., 290(60,61), *301*
Zarraga, M. N., *208, 212*
Zierep, J., *181, 183, 203, 210, 212*
Zilliken, F., *153, 175*

SUBJECT INDEX

Aggregates, particle, 137, 140, 141, 142
Anisotropic molecular fluids, depolarized light scattering by, 55-61
Aspect ratio, 185, 194, 196, 200
Astronomy, 115
Atomic units, 217
Averages, ensemble, 117, 131, 138, 140
  time, 110, 137, 138, 140

Basis functions, 219, 243
Benard convection, 177
  in the atmosphere, 178
  balance theorem, 202
  in bounded containers, 178, 202
  bounded laterally, 196
  with centrifugal force, 208
  with coriolis force, 207, 208
  finite amplitude theory, 200, 201
  heat transfer by, 185, 189-191, 194, 201, 202, 204, 205, 207-209
  higher modes of, 184
  internal heating, 208, 209
  maximal heat transfer, 201, 203
  nonuniform heating, 207
  with rotation, 207, 208
  with shear, 187, 206
  supercritical, 202
  in two component fluids, 178
  wave length, 185, 187, 190, 193, 194, 197-199, 202, 203, 207-209
  wave number, 178, 180, 193
Binary (stars), 109, 116, 137, 140; *see also* Pair, Keplerian
Binary dynamical formation of, 116, 132, 133, 134, 137, 138, 143
Blood coagulation, 145
Born-Oppenheimer approximation, 217
  in quantum hechanical description of light scattering by simple fluids, 12
  separation, 12
Boundary conditions, 119, 121, 122, 142
Boussinesq approximation, 179, 194
Box, 118, 120, 122, 123
  cold, 119
  see also Enclosure
Brillouin's theorem, 248

Calculation, 129
  variational, 129, 132, 134
  see also n-body calculation
Catastrophe, 134
  gravothermal, 115, 122, 124
Cells, 181, 190
  hexagonal, 183, 186-189, 190, 192, 193, 201, 204, 206, 208, 209
  hexagonal patterns, 193, 194, 208
  rectangular, 181
  ring, 181, 183, 187, 194, 203
  square, 181, 200, 201, 205
  supercritical R, 203
  triangular, 181
Classical description of light scattering by simple fluids, 24
  dipole-induced-dipole polarization, 25-27
  iterative solution of effective field equations, 26-27
Cluster, 109, 110, 112-114, 119-122, 124, 126, 133, 134, 142, 143
  of galaxies, 108, 138
  star, 113, 115, 138, 140-141
Cluster expansions, 22, 55
  for depolarized light scattering, 22
  for polarizability of simple fluids, 55
    ab initio calculations, 55, 73, 74
    approximations for triplet component, 69-70
    for dielectric functions, 66
    for Kerr effect, 63
    superposition, 66, 67
Collapse, 123, 124
Collisions, 108, 111, 112, 142
Complement, 164
  cascade, 164
Configuration, 230
  interaction, 217

state functions, 229
Contracted basis functions, 245
Control experiments, 185
Constraint, 129, 132
Cool, 121, 123
Cooperative effects, 121
Core-Halo structure, 122, 123
Correlations, 114, 124-6, 129-134, 141, 142; *see also* Energy, correlational
Coupling, 110
Crossing-time, 110, 117, 119, 120; *see also* Time scales, dynamical
Cutoff, 111, 136, 139, 140, 142; *see also* Volume, cutoff

Density kernel, first and second order reduced, 222, 225, 237
Depolarized light scattering, data, analysis of, 36, 43, 51
density expansions, 43-49
three (and many) body structure 45-48
two body effect, 43-45
isolated binary collision model, 36, 40
gases, 36-40
liquids, 40-43
lattice occupancy models, 50-51
molecular dynamics computations, 51-55
second order Rowan theory, 49-50
Depolarized scattering near the gas-liquid critical point, 80, 81
angle dependence of cross-section, 87, 88
double (versus collision induced) scattering, 78, 80
long range contributions to two, three and four particle correlation effects, 81-87
Depolarized scattering of light by simple fluids, 2, 9, 30
classical description, 34
relationship to binary collision models, 34-36
experimental geometry review of present theories, 3, 6-8
fluctuating local field analysis, 30-33
quantum mechanical descriptions, 9-11, 16-23
binary collisions, 17-19
higher order effects and liquids, 21-23
three body dynamics, 19-21
Dielectric function for simple gases, 65-78
corrections to DID (Kirkwood) theory, 70-77
density expansion, 66
dipole-induced dipole (DID) limit of "exact" theory, 68, 69
Dimensions system, 111
Dynamics, stellar, 108, 109, 112, 115, 117, 125, 126, 129, 142; *see also* Stellar dynamic systems

Enclosure, 117, 119, 120, 122
adiabatic, 118
isothermal, 118, 121, 123, 124
specularly reflecting, 118, 121, 123
Encounters, binary, 111, 123, 133
Energy, 112, 126, 143
conservation of, 108, 118, 122, 131, 134
correlational, 131, 132, 143
internal, 115, 129, 130, 131
kinetic, 110, 118, 119, 123, 126, 130, 133, 138, 139
potential, 110, 118, 122, 126, 131, 133, 135, 136, 138, 141-143
total, 119, 126, 129, 132, 134, 135, 137, 139, 141-143
Ensemble, 112, 114, 134, 141, 142
canonical, 112, 134
grand canonical, 112, 134
microcanonical, 113, 115, 134-142
Entropy, 108, 127, 130, 139, 140, 142
information-theoretic, 127
conditional, 128
thermodynamic, 130
Equation, 133
Liouville, 133
n-body, 110, 113, 127, 128, 134
violations of, 116, 123, 124, 129, 132
of motion, 110, 115, 128, 129, 133, 135, 142
Vlasov, 133
Equilibrium, 108, 113, 115, 124, 130, 132, 134, 138, 140-143
Maxwellian, 131

metastable, 113, 114, 134, 141
Ergodic, 112, 113, 138, 140
Evolution, 113, 128, 129, 132, 133, 134, 141-143; *see also* Time scales, evolution
Evolutionary tree, 162
Exclusion, 124, 126
Experiments, 120, 122-124, 126
numerical, 115, 125, 133, 134, 142, 143
thermodynamic, 116, 118

Factors, blood, 152
Fibrinogen, 158
Finite system, 112, 126, 142
Fluctuations, 121
Force, inverse square-law, 108, 110, 111, 116, 124
long range, 108, 112, 114, 116, 134
see also Long-range
Frozen core approximation, 240, 253
Function distribution, 117, 126, 131, 133
conditional, 128
frequency, 123
bivariate normal, 129
conditional probability, 129
marginal, 129
multivariate Gaussian, 130
see also Probability
H, 130
Boltzmann, H_B, 127-134, 141-143
Gibbs H_G, 127-129, 132, 141, 142
Rate of increase of H_B, 143
homogeneous, 135, 136, 141
n-particle, 125, 127, 132, 142
as product of single-particle, 127
pair configuration, 131
reduced, 125, 127, 132, 141
single particle, 125, 127, 129, 131-133
see also Probability

Galaxy, 108, 109, 115, 134
Gaussian-lobe functions, 220
Gaussian-type functions, 220
Gravothermal, 115, 122, 124

H-theorem, 113, 126-134, 141-143
Hamiltonian, 108, 123, 124
Hartree-Fock Equations, 225, 227, 233
Hartree-Fock model, 224
Heat bath, 124, 134
Heat capacity, 118
Hierarchy, 109, 115
BBGKY, 125
Hot, 121, 122, 123
Housholder tridiagonalization, 239

Inequalities, 127, 128, 130, 132
Infinitesimal disturbances, 178, 209
Infinite system, 112, 126, 132
Information, 142
theory, 127
Instability, 111
numerical, 117, 138
see also Jeans instability
Integrals, 113
first, 113, 129
of motion, 113
Interaction, 110, 111, 116, 122, 124, 132, 138
graviatational, 110, 112
Isothermal, 123
structure, 123

Jeans instability, 111, 112, 126

Kerr effect for simple fluids, 61-65
Kinin System, 167
Koopmans' theorem, 249

Lagrangian multipliers, 222, 236
Limit, thermodynamic, 112, 117
high-low density, 117
see also Thermodynamics
Liouville, theorem, 113, 123, 124, 127, 130, 132
Long-range, 108, 111, 124; *see also* Force, long range
Long-time, 112, 114; *see also* Memory, ling time

Marangoni number, 192, 193
Marginal curve, 180
Marginal stability, 179, 180
Matrix, 130
determinant of, 130
MCSCF Equations, 222, 223
Mean field, 111, 114
Memory, long-time, 112, 142
Michaeles-Menton model, 149
Model potential, 242

Molecular wave functions, 247
Momentum, 118
  angular, 118, 122, 129, 135
Multiconfiguration self-consistent field method, 221
MVDVG equations, 222, 223

Navier-Stokes equation, 178
n-body calculation, 116, 117, 119, 123, 134, 138
n-body problem, gravitational, 108, 137-139, 141
  systems, 109, 116, 117, 132

Observation, 114, 133, 134, 137, 143
Optimized valence configuration method, 253
Overstability, 179, 180, 207

Pair, Keplerian, 116, 125, 126, 132; *see also* binary (stars)
Particles, 108
Platelets, 157
Polarizability distortion in simple fluids, 34
  Mazur theory, 34-35
Polarized scattering of light by simple fluids, 4, 15
  classical description, 27, 28, 34
  quantum mechanical descriptions, 9, 15, 16
  references to recent literature, 4
    Einstein relations, 16
    lineshape, 15
Probability, 137, 138
  conditional, 128, 129, 141
  see also Function, distribution; Function frequency
Pseudopotentials, 240

QL algorithm, 239

Random phase approximation, 251
Rayleigh number, 178, 190, 191
  critical, 178, 180, 184, 185, 188-191, 193-197, 199, 202-204, 207, 208
  high Rayleigh number, 209
  supercritical, 209
Relaxation, 111
Renin-angiotension system, 166
Reversibility, 113
Rolls, convective, 181, 183, 187-191, 194, 210, 205, 209
  circular, concentric, 194-197, 204, 206-208
  counter rolls, 207

SCF Xα SW (self-consistent-field Xα scattered wave method), 258
Screening, 108, 111, 112
Second harmonic generation by simple fluids, 23, 24
Self-gravitating systems, 109, 110, 115, 116, 126, 132, 139
Simple fluids, definition of, 2
Slater determinant, 218
Slater-type functions, 220
Space, configuration, 124, 131, 136, 137, 141
Space phase, 124, 126, 138
  division into cells, 126
  μ-space, 113, 124-126, 133
  Γ-space, 123, 124, 133
  see also Volume, phase
Specific heat, 139, 140
  negative specific heat, 115, 122, 139
Spin orbital, 219
Stability, 113, 133, 134, 143
  theory, linear, 178
    nonlinear, 200, 203, 208
  see also Jeans instability
Statistical mechanics, 108, 109, 111-115, 127, 139, 142
Stellar dynamic systems, 108, 109, 113, 114, 124, 126, 127, 131, 134
Stochastic process, 112
  Markov, 112, 142
Subclusters, 119, 122
Subcritical instabilities, 188, 205, 208
Subdivisions, 112, 134, 139
Subgroup, 141
Surface, energy, 136, 137, 139
  tension, 179, 187, 192, 194
System, 117
  hard-sphere, 117
  Maxwellian, 130
  statistical, 117
  stellar, 109, 111, 112, 116, 121, 124, 126, 130-134, 142

Taylor number, 207
Temperature, 118, 120-124, 139
  of box, 119, 120
  thermodynamic, 116, 124
Thermal noise, 185, 206
Thermodynamics, 112, 115-118, 130, 139
  analogies, 139, 140
  second law of, 139
Time-scales, 110, 113-115, 124, 133, 141
  dynamical, 110, 123, 124, 133
  evolution, 113
  secular-relaxation, 111, 113, 123, 124, 133, 134, 141
Transfer, of energy, 119, 121-124
  rate of, 119-121, 124
Transition density kernel and matrix, 222

Uncorrelated state, 129
Universe, age of, 111, 113, 143
Unstable, 111, 117, 143

Virial ratio, 134, 138
Virtual orbitals, 250
Volume, infinite, 112
  cutoff, 139, 140
  phase, 115, 123, 124, 135-137, 139-142

Xα, statistical exchange approximation, 258

Zero differential overlap approximation, 255

Date Due

TRENT UNIVERSITY
0 1164 0085788 8

QD453 .A27 v.26 1974

Advances in chemical physics

DATE | ISSUED TO

223834
223834

ISBN 0-471-60931-4